EVALUATION OF THE

MADHYA PRADESH

SAFE CITIES INITIATIVE

2013-16

Endline Report

Produced for: DFID India

August 2016

Authors:
Jenny Holden
Macartan Humphreys
Sohail Husain
Sarah Khan
Summer Lindsey

With contributions from Sally Neville, Lyndsay McLean, Kalpana Viswanath, Jessica Jacobsen and Sarah Donachie.

new concept
creativity & collective work transforms

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Social Development Direct

Contents

Tables

Figures:

Annexes:

Acknowledgements

The impact evaluation of the Madhya Pradesh Safe Cities Initiative (SCI) and the production of this report have only been possible because of the valuable contributions from a broad range of individuals and organisations. We would like to thank everyone who has been involved.

We owe particular thanks to those who participated in data collection: respondents to an individual panel survey and participants in the focus group discussions (FGDs) and key informant interviews (KIIs). Without their generosity in terms of the time they gave, and their willingness to be asked questions on a sensitive subject, this evaluation would not be possible. We also thank the large team of enumerators and supervisors from NCIS who worked long hours to conduct the quantitative survey.

We are grateful for the support and cooperation of the Urban Development and Environment Department (UDED) within the Government of Madhya Pradesh (GoMP), in particular Mr Sanjay Shukla IAS, Commissioner Urban Administration and Development and Project Director for Madhya Pradesh Urban Infrastructure Investment Programme (MPUIIP), Mr C.U. Roy, Deputy Director (Planning) MPUIIP and the cluster teams and Nodal Officers in the Municipal Corporations in Bhopal, Gwalior, Indore and Jabalpur. We also thank the implementation partners at GHK IPE Global who delivered the SCI, in particular, Richard Slater, Manisha Telang, Asmita Basu and Shahzad Khan.

We would like to thank Jagori - a women's rights organisation based in New Delhi– who led the qualitative fieldwork and also provided specialist training to the quantitative field teams on how to manage the sensitivities associated with conducting VAWG research. A team from Jagori including Sarita Baloni, Chaitali Haldar, Praveen Verma, Mayur Chetia, Shruti Batra, Shreya Banerjee, Surabhi Mehrota, Geetha Nambisan, and Madhu Bala undertook the qualitative fieldwork at endline. We would like to give special thanks to Surabhi Mehrota who managed the qualitative research, and the Director of Jagori, Geetha Nambisan, who provided excellent leadership and oversight to ensure smooth running of the process.

We would like to thanks Virginia Morrow at Oxford University and Mary Ellsberg at The George Washington University for their advice and wisdom on the ethical considerations we needed to respond to in relation to this work. Also to Joe Holden for his support in defining our approach to analysing value for money in relation to the SCI. The invaluable learning from each of these contributions has significantly enriched our evaluation approach, methods and analysis.

We are grateful to the members of the peer review group for this evaluation who helped to shape the evaluation design and methods - including Zoe Stephenson, Shireen Jejeebhoy, Sambit Rath and Rajib Acharya. We also benefited from discussions with leading experts in the field including Suneeta Dhar from Jagori, Nandita Bhatla from ICRW, Shireen Jejeebhoy from the Population Council, participants at EGAP 14 who gave feedback on the design, and Suvakanta Roy Swain from FHI 360 who verified our endline outcome measures and generously shared their experiences of conducting VAWG research in India.

Finally, we thank the Department for International Development (DFID)-India for commissioning this work, with the intention of making a valuable contribution to the evidence base about what works to prevent violence against women. Our particular thanks go to Arundhuti Roy Choudhury.

Evaluation team

Social Development Direct

Social Development Direct (SDDirect) is the UK's leading niche provider of innovative and expert social development assistance and research services. We provide technical specialists who offer high quality support and advice to leading international development agencies including the United Nations (UN) agencies, DFID and international non-governmental organisations (iNGOs). We offer in-depth technical expertise on a core set of social development issues, and provide a wide range of services including technical support and advice to programme design, planning, strategy and management; programme reviews, monitoring and evaluation; and qualitative and quantitative analysis. We have a large gender portfolio and work extensively on violence prevention and response including leading the Violence Against Women and Girls (VAWG) Helpdesk, a DFID-funded global service that provides research and short-term technical assistance on the full range of violence issues. We are also a core research partner implementing DFID's flagship 'What Works' to Prevent VAWG Programme.

SDDirect senior consultants Sally Neville and Jenny Holden were the Team Leaders for the evaluation at baseline and endline respectively, working alongside Evaluation Director Sohail Husain, VAWG Expert and Qualitative Research Advisor Lyndsay McLean, Safe Cities Expert Kalpana Viswanath, and SDDirect consultant Jessica Jacobson and Project Officer Sarah Donachie.

Columbia University

Professor Macartan Humphreys, Sarah Khan and Summer Lindsey of Columbia University's Center for the Study of Development Strategies (CSDS) led the quantitative and experimental components of the study. CSDS focuses on the design and implementation of Randomised Control Trials (RCTs) and analysis of experimental data, and has supported projects examining development interventions in a wide range of settings including Uganda, Congo, Liberia and Indonesia.

New Concept Information Systems

New Concept Information Systems (NCIS) is an Indian consultancy that specialises in social development. It brings particular expertise in social evidence, research, impact assessment and monitoring and evaluation. They have over 25 years' experience of working in the development sector with central and state governments, UN agencies, non-government organisations (NGOs) and the corporate sector. They have worked for 100+ clients across India, including the Government of India, state governments, UN agencies, multi-lateral and bi-lateral donors, institutions and corporations. New Concept has delivered over 200 evaluation projects including the baseline and endline studies for the SCI in Madhya Pradesh. The team leader for the survey was Sanjay Tiwari, working with a team of Field Supervisors: Subhash Sinha, Rajeev Singh, Krishna Negi, Sunil Bhaduriya, Anjali Mishra, Preeti Mishra; and two ICT Programmers: Amardeep Ujjain and Dhiraj Kishore. Founding Director of New Concept, Vimala Ramakrishnan was the senior resource person.

The content of this report and all opinions expressed are the responsibility of the authors.

List of abbreviations

BPL	Below Poverty Line
CCT	City Cluster Team
CSO	Civil Society Organisation
CSA	Community Support Agency
DFID	Department for International Development
FD	Female Direct (Beneficiary)
FGD	Focus Group Discussion
FI	Female Indirect (Beneficiary)
GoMP	Government of Madhya Pradesh
ICC	Intra-Cluster Correlation
IP	Implementing Partner (GHK/IPE Global)
IPV	Intimate Partner Violence
KII	Key Informant Interview
LSHTM	London School of Hygiene and Tropical Medicine
MD	Male Direct (Beneficiary)
MDE	Minimum Detectable Effect
MFIG	Micro-Finance Institution Group
MI	Male Indirect (Beneficiary)
MIO	Male Indirect Older (Beneficiary)
MIY	Male Indirect Younger (Beneficiary)
MP	Madhya Pradesh
MPUIIP	Madhya Pradesh Urban Infrastructure Investment Programme
MPUSP	Madhya Pradesh Urban Services Programme
NCIS	New Concept Information Systems
NFHS	National Family Health Survey
NGO	Non-Governmental Organisation
OBC	Other Backward Castes
PDA	Personal Digital Assistant (Device)
PVAW	VAW in public spaces
RAY	Rajiv Awas Yojna (National Housing Scheme for the Urban Poor)
RCT	Randomised Controlled Trial
SC	Scheduled Caste
SCI	Safe Cities Initiative
SD	Standard Deviation
SHG	Self Help Group
ST	Scheduled Tribe
ToC	Theory of Change
UDED	Urban Development and Environment Department (UDED)

ULB	Urban Local Body
VAW	Violence Against Women
VAWG	Violence Against Women and Girls
WSA	Women's Safety Audit

Box 1: Explanation of the use of 'VAW' or 'VAWG' in the report

Throughout this report there are references to both 'VAW' (violence against women) and 'VAWG' (violence against women and girls). This is based on a deliberate effort to make a distinction between the two. The overall aim of the Safe Cities Initiative is to reduce VAWG through engagement with groups of boys and young men and with women in SHGs. Girls are not directly targeted through the Programme but it is intended that they will benefit indirectly from the interventions. The evaluation does not, however, involve data collection from girls. This means that many of the endline findings relate specifically to VAW as only women's views and experiences were directly captured. However, the qualitative data does present views that were expressed more broadly on VAWG, as FGD participants often gave their perceptions on violence experienced by girls in their community. The Programme sought to encourage women, men and boys to take action to prevent and respond to VAWG in their communities.

Executive summary

Globally, violence against women and girls (VAWG) is one of the most widespread forms of abuse - affecting on average one in three women in their lifetime.[1] India is no exception, where data from the 2005-2006 National Family Health Survey (NFHS-3) found that over one third of married women have experienced physical or sexual violence by their husband.[2] In Madhya Pradesh, the same survey found almost half of all women (46%) have experienced physical or sexual violence by an intimate partner.

In response to limited evidence for the effectiveness of violence prevention interventions in India and the general lack of evaluations that capture impact, in 2013 DFID India commissioned **Social Development Direct (SDDirect),** together with **Columbia University**, and **New Concept Information Systems (NCIS)** to conduct an independent impact evaluation of the **DFID-funded Safe Cities Initiative (SCI) in Madhya Pradesh (2013-2015)** – to provide robust evidence on what interventions work – and do not work - to reduce VAWG. This report presents results of the evaluation including insights for future VAWG focused programmes and evaluations.

The Programme

A component of the DFID-funded Madhya Pradesh Urban Infrastructure Investment Program (MPUIIP), the SCI is quite unique in VAWG prevention programming, as it aimed to address *both* **intimate partner violence (IPV)** and **violence and harassment against women and girls in public spaces.** The programme also represents a commendable attempt by DFID and the Government of Madhya Pradesh (GoMP) to scale up and replicate promising approaches to VAWG prevention in 250 urban slums in four cities (Bhopal, Indore, Jabalpur and Gwalior).

The SCI was delivered through a partnership comprising the Government of Madhya Pradesh (GoMP) and the implementing partner (IP), GHK/IPE Global. The interventions were delivered by four community support agencies (CSAs). The SCI (the 'Programme') included the following three interventions:

- **T1: Self-help group (SHG) strengthening module** – which aimed to strengthen existing and nascent women's SHGs.
- **T2: SHG Strengthening +VAW module** – which built on the first intervention and aimed to increase women SHG members' awareness of VAWG and capacity to take actions to address it.
- **T3: Life skills module with men and boys** – which aimed to increase men and boys understanding of underlying causes of VAWG and to build their capacity to challenge harmful social norms and take actions against VAWG at the community level.

The programme aimed to impact both **direct beneficiaries** (members of SHGs and men and boys groups) and **indirect beneficiaries** (members of the wider community within the slum).

Evaluation methodology

The evaluation was designed to identify which of the Programme's interventions (if any) are most effective – and why. This is reflected in the selected design for the evaluation: **a factorial Randomised Controlled Trial (RCT),** with treatment effects estimated through a **mixed-methods data collection** approach at baseline and endline. The quantitative component included an **individual panel survey** of approximately 7,500 individuals. The **qualitative component** included 48 focus group discussions (FGDs) and 19 key informant interviews at endline. The outcome areas of focus include both **primary**

[1] WHO, South African MRC and LSHTM (2013)

[2] India Ministry of Health and Family Welfare (2006).

outcomes on prevalence of VAW as well as a set of **intermediate outcomes** along intended pathways to reducing VAW including individual attitudes and norms, women's economic empowerment, community action against VAWG and reporting of VAWG. The panel survey also included a number of **innovative measurement techniques** such as an embedded list experiment designed to address issues of underreporting of sensitive issues such as VAWG[3], behavioural measures to capture people's willingness to take public action, and measures to capture social norms.

Findings on Impact

IPV and public VAWG: There is no evidence that the SCI led to a reduction of either IPV or violence and harassment against women in public spaces. The qualitative component indicates that some SHG members may face an initial increase in IPV, which later tends to decline as husbands become aware of the benefits of membership. However, this finding is not supported by the quantitative evidence.

Attitudes and norms: There is limited evidence that the SCI led to an improvement in attitudes. Notably, there is no evidence of change in gender equitable attitudes or individual attitudes towards IPV or VAWG in public spaces. However, the qualitative data indicates that some male direct beneficiaries (MDs) of the Life Skills intervention (T3) may be beginning to challenge harmful gender roles particularly around unequal roles in the household. There is limited evidence of treatment effects on social norms, with some evidence of weak effects but not in a way consistent with Programme goals. The findings indicate that descriptive norms – i.e. what people think other people do, may be a more powerful driver of VAWG than prescriptive norms – i.e. what people think other people expect them to do, in urban slums in Madhya Pradesh. Consequently, despite individual attitudes against IPV and violence and harassment in public spaces, they remain normalised behaviours. Moreover, the qualitative data highlights harmful social norms which continue to perpetuate VAWG including prevailing norms that IPV is a family matter, it is a wife's obligation to have sex with her husband even if she doesn't feel like it, and survivors are to blame for violence perpetrated against them.

Reporting VAW: There is some evidence of improvements in female direct beneficiaries' (FDs) individual attitudes towards reporting IPV among SHG members who received the SHG+VAW module (T2). However, there is no such improvement towards attitudes towards reporting VAWG in public spaces. Despite positive attitudes towards reporting, respondents expect only 20% of women will report IPV. There is limited and inconsistent evidence of treatment effects on norms around reporting, but no evidence of impact on reporting VAWG to police or protection officers. The qualitative data highlight significant barriers towards reporting including persistent norms, fear of social sanctions and low trust in police and local response services. However, there is some qualitative evidence that the SHG+VAW module may have increased women's awareness of reporting mechanisms as well as women's willingness to talk about VAWG. Although this change is not yet associated with an increase in reporting – it may be an important first step to achieving social change.

Women's economic empowerment: There is no evidence that the SHG interventions led to an increase in women's income or increased financial independence. In fact, there is some evidence that the SHG strengthening module (T1) may have led to worsening of control over income for FDs in T1 slums, potentially signalling backlash from husbands. Moreover, the qualitative component highlights unemployment and lack of viable livelihood activities as significant challenges experienced by SHG members. However, the qualitative component also indicates that SHG membership may have

[3] List experiments involve providing respondents with a predefined list of behaviours and asks them to report *how many* behaviours from the list they have encountered, without specifying which ones. Such list experiments have been successfully used in previous studies to gather information on respondents' attitudes and behaviours which they do not readily admit to, for example racist or extremist views.[3]

broader benefits for women's empowerment including improved social networks and confidence – which in future may reduce vulnerability towards VAWG.

Women's mobility and feelings of safety: There is no evidence of change in women's mobility or feelings of safety. Women continue to sometimes feel unsafe in their slum – particularly at night time. And women's mobility continues to be severely restricted. Whilst the qualitative data indicates SHG members perceive that their own mobility has improved, these improvements are limited to movement for SHG meetings with no evidence of wider improvements in women's mobility as a result of interventions.

Actions taken to address VAWG: There is strong evidence of treatment effects on FDs in T2 slums expectations of receiving support from SHGs. However, there is no similar evidence of changes in expectations of receiving support from the police. There is no evidence of effects on women or men's actions taken to prevent or respond to VAWG or their willingness to engage others on the issue.

Heterogeneous Effects: In annexes we provide extensive analysis of heterogeneous effects. On some factors – such as city – there is evidence of effect heterogeneity though not consistent i.e. suggestive of positive effects for some groups and negative effects for other groups. There is no city or other subgroup for which we consistently find positive support across outcomes.

Conclusions

The factorial design allowed us to assess a variety of programme types and combinations. The quantitative component of this evaluation finds little evidence that any of the treatments included in the SCI or combinations of them had positive impacts on the hypothesised outcomes.

Note that the quantitative analysis examined the effects of the Programme on 61 outcome variables for between two and four different populations (including both direct and indirect, male and female beneficiaries) drawn from each slum, resulting in a large number of statistical tests. Thus, when analysing the effectiveness of the various treatment arms, it is important to focus on the consistency of findings across outcomes rather than on any particular result.

Overall, there are a very small number of Programme effects on particular outcomes, and over half of the effects are not in the hypothesised direction. Furthermore, results are most limited where we would most expect to see them – at the level of the direct beneficiaries.

Findings from the qualitative analysis on outcomes of interest highlight wider perceived benefits of SHG membership including improved social networks and confidence as well as some signs that key messages in the Life Skills module may have been effective in encouraging men and boys to challenge unequal gender roles in the household. However, there is limited evidence in wider shifts in attitudes, norms and behaviours from the qualitative data, which is broadly supportive of the quantitative findings. In particular, the qualitative data highlights persistent harmful norms and attitudes which continue to drive and sustain VAWG in urban slums in Madhya Pradesh.

Learning from the null results

In order to account for the evaluation findings, and to generate learning and insights to inform future VAWG programmes and evaluations, we explored three possible explanations for the null results. While the quantitative component of this study was not designed to differentiate between these three possibilities, we provide information based on monitoring data, qualitative data as well as the accounts provided by implementers

1. **Implementation failure -** i.e. Due to challenges in implementation, the SCI was not implemented as intended, and there was a lack of Programme fidelity. This explanation

suggests that underlying issues were not associated with application specific to this context, but simply an issue of delivery.

2. **Theory failure** – i.e. The theory of change associated with this programme does not reflect how change happens on the ground. This means that the Programme would not have worked even if implemented perfectly.
3. **Measurement error** - i.e. The research failed to reveal the true results, because the design, the measurement or the analysis was weak.

There is available evidence for **implementation weaknesses** which supports a lack of implementation fidelity. In particular, it is likely that with significant delays to implementation, the Programme duration may simply have been too short to expect measurable changes to occur in deep seated attitudes, norms and behaviour. Furthermore, weak adherence to the intervention design in terms of dosage, quality and intensity of delivery likely affected programme outcomes.

There also appears to have been wide variation in how the Programme was implemented. Monitoring data and IP accounts suggest variation in implementation quality by city. In the quantitative heterogeneous effects analysis by city, our findings do lend some support to the idea that there were differences in treatment effects across cities (in line with anecdotal evidence from the IP) with, marginally more positive effects in Bhopal and more negative effects in Gwalior. However, these differences are not necessarily due to differences in implementation as they may also reflect differences in local context.

There is also some evidence to suggest **weaknesses in the underlying Programme theory.** In particular, the findings question the SHG model as an effective way to economically empower women, with suggestive evidence that membership may lead to reduced control of income. Furthermore, the results suggest that focusing on awareness of rights and laws (key approaches of the SHG+VAW and Life Skills modules) is not sufficient to shift social motivations of human behaviour. Furthermore, Programme strategies for engaging with local response mechanisms may not have been sufficient, and consequently low trust in police response and perceived inadequacies in local response mechanisms continue to create significant barriers for reporting IPV and pubic VAWG, whilst supporting norms around perceived impunity for perpetrators.

A third possibility for lack of effects may also be due, in principle, to **weak measurement**. The measures used to assess prevalence of VAW were based closely on international standard measures. However the baseline/endline comparisons suggest that these standard measures are very 'noisy'. These measures were complemented by a set of indirect experimental measures and by a set of behavioural measures. These each come with specific advantages. They provide some encouraging validation results for the indirect measures when compared to more direct measures. The behavioural measures have the advantage of measuring in a very direct way a set of intermediary outcomes of interest—the ability of groups to mobilise to bring about change. Due to the study's adherence to standard measures along with added measurement innovations, if null findings stem from measurement, these issues are also relevant to VAWG studies more broadly.

In sum, while a central goal of this evaluation was to understand the effectiveness of specific VAWG prevention programmes in bringing about change, questions about measurement and implementation make it difficult to pinpoint whether null results are due to the ToC, implementation, or even measurement.

From analysis of the null results of this evaluation, we suggest two important insights for evaluations of VAWG programmes:

- **Importance of monitoring data to measure implementation fidelity.** Monitoring data is just as important as evaluation data and is key to tracking implementation fidelity. Evaluations of

this kind would benefit from rigorous quantitative monitoring data that can be included in quantitative analyses at the level of the randomisation. More thorough qualitative evidence such as beneficiary feedback, observations of the quality of delivery, and monitoring of signs of backlash would also help with interpretation of findings.

- **Robustness of VAWG measures.** As part of the evaluation, we were able to assess the reliability of commonly used VAWG prevalence measures by both comparing consistency over time and by comparing direct measures with indirect measures. Both analyses highlight weaknesses in the standard sensitive measures commonly used in evaluations of this form, with potential wider implications for evaluations in this area.

Insights for future VAWG programmes and evaluations

While the statistical analysis provides no clear explanation for the results, by comparing what we know about the SCI with evidence and experience in the wider VAWG field, SDDirect offers the following practical insights for future VAWG prevention programming:

- **Achieving transformational change at scale may require higher intensity and consistent programming.** Our results highlight challenges in ensuring quality of scaled up VAWG component programmes. Our conclusions highlight the relative importance of key implementation weaknesses in accounting for the disappointing results, and underscore the importance of quality and intensity of delivery to achieving sustained impact on attitudes, norms and behaviour. In future designs of scale up programmes, the recommendation is to more thoroughly unify processes, delivery, and implementation to hedge against variation especially if the goal is to examine overall (rather than area-specific) programme effects.
- **Achieving change in VAWG is likely a long-term process, requiring sustained resources and realistic timeframes.** Programmes with long-term investment, sustained resources and realistic timeframes are likely to be more effective and to have both greater and more sustained impact.
- ***Who* delivers may be just as important as *what* is delivered.** Our findings highlight the importance of ensuring partner agencies have sufficient VAWG expertise and that field staff have the capacity and support in VAWG prevention programming to model gender-responsive behaviours and motivate and support communities to tackle VAWG.
- **Shifting VAWG likely involves tackling deep seated harmful social norms.** Our findings highlight the pervasive nature of harmful social norms around IPV and public VAWG in the target areas. Furthermore, the findings suggest that in the context of urban slums in Madhya Pradesh descriptive norms – i.e. what people think other people do – may be a more powerful driver of VAWG than prescriptive norms – what people think other people expect them to do. In this context awareness raising campaigns at worst are unlikely to be effective, and at best may exacerbate these norms. Whilst there is no 'one size fits all' approach to shifting social norms, emerging evidence suggests efforts need to be made to diagnose and address specific norms at play in different contexts.
- **Challenges of external validity of adapted programmes.** Lessons from this evaluation suggest that experimental programmes of this nature which are based on evidence from other contexts, require a period of adaptation and learning to adapt approaches prior to implementation in order to ensure effectiveness of the approach and underlying theory.
- **Ineffective local VAWG response may significantly limit effectiveness of VAWG prevention.** Wherever possible, prevention programmes should engage in strengthening VAWG response as a key strategy to ensure sustained social change and adhere to do no harm principles. For programmes which engage in providing local response and mediation, women's safety must always remain the priority, and it is important to ensure that those who are mediating have sufficient skills and are provided with appropriate training and support.

PART A: BACKGROUND AND METHODOLOGY

1. Background

Introduction

In cities, such as those in Madhya Pradesh, and other urban areas in India, women and girls may be subject to violence in many forms in both private and public spaces– including the home, the neighbourhood, at workplaces and on public transport. This exposure to, and risk of, violence impairs their health, restricts their mobility and reduces their access to the opportunities cities offer, including education, employment, recreation and political participation. The cumulative effect is to significantly limit the freedom of women and girls to exercise their rights as equal citizens. VAWG also has adverse impacts on the economic and social development of cities.[4]

Preventing VAWG is a top priority for the UK Government, which is being addressed internationally through various ministries including the Department for International Development (DFID). In response to limited evidence for the effectiveness of violence prevention interventions in India and the general lack of evaluations that capture impact, in 2013 DFID India commissioned Social Development Direct (SDDirect), together with Columbia University and New Concept Information Systems (NCIS) to conduct an impact evaluation of the DFID-funded **Safe Cities Initiative (SCI)** in Madhya Pradesh – to provide robust evidence on what interventions work – and do not work - to reduce VAWG. The impact evaluation of the SCI is part of a wider commitment to identify interventions that can measurably reduce the incidence of VAWG.[5]

In particular, the impact evaluation of the SCI explores two types of interventions and their impact on VAWG:

- **Women's Self Help Groups (SHGs):** A popular development intervention to address gender inequality and 'empower' women. These groups may have economic, legal, health and/or cultural objectives, often with the primary aim of empowerment – typically involving savings, credit or social involvement as instruments of empowerment.[6] India has a long history of SHG activity, including the SHG-Bank Linkage Programme, which was initiated in 1992 and has been expanding ever since. The impact of SHGs on VAWG and their potential as a 'vehicle' for addressing VAWG are under-researched topics and so the evidence base is limited, and available findings relating to impact on intimate partner violence (IPV) are mixed.[7] This rigorous impact evaluation of the SCI aimed to fill this knowledge gap.[8]
- **Life Skills work with men and boys:** The importance of working with men and boys in the prevention of VAWG is now widely accepted by policy makers and experts.[9] However, the

[4] See DFID What Works programme component 3 http://www.whatworks.co.za/about/what-works-components/what-works-economic-and-social-costs-of-vawg

[5] DFID India is also funding an impact evaluation of a VAWG prevention programme in Bihar, India being undertaken by the Population Council.

[6] Brody et al (2016)

[7] A 2009 systematic review of economic empowerment and IPV found that, whilst economic development and poverty reduction may offer protective factors, context specific factors influence whether financial autonomy is protective or associated with increased risk. Vyas, S.; Watts, C (2009)

[8] Brody et al (2016) conclude from their systematic review of SHGs on empowerment that more rigorous evidence of the impact on IPV is needed.

[9] Jewkes et al (2014)

building of an evidence base of effective interventions is at an early stage in both scope and scale, with rigorous evaluations of interventions often missing, methodologically limited or inconclusive.[10] In particular, a recent review found limited evidence of effects on social norms, including transforming harmful masculinities; a lack of evidence from the Global South; and a lack of data on optimal dosage[11], scope and scaling up of interventions.[12]

This report presents results of the impact evaluation.

Purpose of the evaluation

In June 2013, SDDirect was commissioned to lead an independent impact evaluation of the SCI in partnership with Columbia University in New York and NCIS in New Delhi. The evaluation design seeks to contribute to both Programme accountability and Programme learning, and has a dual purpose**:**

- **To conduct a scientifically robust impact evaluation**, which will assess the achievement of key results and the extent to which these can be attributed to Programme interventions;
- **To generate learning and insights** into complex processes of change, what works or doesn't work and why, in order to inform on-going and future Programme development.

The evaluation has therefore been designed to achieve the following **three objectives**[13]:

1. To assess a number of core Programme results and the extent to which any observed changes are attributable to the Programme interventions;
2. To generate data on a number of intermediate results and indicators which will provide greater insight into the processes of change and answer 'how' and 'why' questions;
3. To make a broad assessment of the relevance, (cost)-effectiveness and sustainability of the Programme.

The evaluation is not intended to simply generate evidence of any effects which can be attributed to the Programme overall, but more specifically to identify which of the Programme's interventions (if any) are most effective – and why. This is reflected in the selected design for the evaluation: a factorial Randomised Controlled Trial (RCT) (outlined in section 2).

Dissemination activities of the evaluation findings at endline will be targeted at audiences in India and internationally in order to further the evidence base of 'what works' in preventing VAWG, with an emphasis on informing future investments in programming.

Structure and content of this report

This report is divided into **four parts**:

Part A provides an introduction to the evaluation providing information on the Programme, evaluation design, sample, approach and methodology.

The primary analysis of evaluation data is presented in **Part B**, beginning with the socio-demographic characteristics of the evaluation sample followed by a brief overview of Programme implementation. We then present the key findings in relation to each of the primary and intermediate outcome measures by intervention type and triangulating with qualitative data where appropriate. Finally we present evidence of any differences in treatment effects by city.

[10] Ricardo et al (2011)

[11] We use dosage throughout the report to refer to the quantity of intervention delivered.

[12] Holden, J (2015)

[13] The terms of reference for the evaluation can be found in annex 1.

Part C concludes with a discussion of the results and explores a number of plausible explanations to account for the findings. **Part D** draws together insights for future programmes and evaluations drawing on wider evidence and experience in the VAWG field.

2. Description of intervention design

Overview of the Safe Cities Initiative

The SCI ('the Programme') is a small additional component of the DFID-funded **Madhya Pradesh Urban Infrastructure Investment Program (MPUIIP)** – which was implemented in the period 2013-2015.[14] The SCI was delivered from 2013-2015 through a partnership comprising the GoMP - including the Urban Development and Environment Department (UDED) and Urban Local Bodies (ULB) - and the implementing partner (IP), GHK/IPE Global. The interventions were delivered by four community support agencies (CSAs).

The SCI reflects an **ambitious and innovative attempt** by DFID and the GoMP **to scale up promising approaches to VAWG prevention** in 250 urban slums across four cities in Madhya Pradesh **(**Bhopal, Gwalior, Indore and Jabalpur).

Unlike many VAWG prevention programmes, the SCI was intended to address **both IPV and violence and harassment against women and girls in public spaces[15]**. It is also unlike many other violence prevention programmes, particularly in India, in that it is focussed on **urban slums** rather than rural communities.

Programme design and intervention packages

The SCI was designed to include activities and interventions at both the slum level and the city/ state level:

- At **slum level** with women's SHGs and with men and boys' groups to raise awareness, challenge gender norms and support efforts to prevent and respond to VAWG. The groups were encouraged to engage with others in their communities in order to promote positive change in terms of attitudes and behaviours;
- At **city and state level** with ULBs and interdepartmental platforms by focusing on improving the capacity of state institutions to respond to and address VAWG, and encourage engagement with communities on violence prevention.

While activities and engagement at the city and state level were intended to strengthen longer-term support and more widespread Government action to prevent and respond to VAWG, it was anticipated that during the Programme timeframe, the slum-level activities would have the greatest impact. Therefore the main focus of the evaluation was on estimating the effects of the following **three slum level intervention modules:**

[14] Since 2006 DFID-India has worked with the Government of Madhya Pradesh (GoMP) to deliver poverty reduction programmes in urban slums.

[15] For ease of reference often referred to as public VAWG throughout this report.

Table 1: Description of Programme Interventions

Intervention	Description
T1: SHG Strengthening Module	This aimed to strengthen existing and nascent women's SHGs.[16] The main activities included training of all SHG members (10-15 per group) in SHG formation, organisation and strengthening; training a selection of SHG members in record and book-keeping; and engaging SHG members in linking with financial institutions and undertaking exposure visits; and regular group meetings. Also covered was basic gender training with information on VAWG referral networks.[17]
T2: SHG Strengthening + VAW Module	This module built on the first intervention and aimed to increase women SHG members' understanding of the root causes and trigger factors related to VAWG. It aimed to build women's capacity to take action and respond to VAWG through trainings, awareness raising, and community mobilisation. It included initial training for selected SHG members, a training for all SHG members, regular monthly meetings and a range of community level activities and events. It also included interface meetings with existing service providers to connect the SHGs and their members with services to prevent and respond to VAWG. Women members were also encouraged to undertake community-level women's safety audits (WSAs) of their slum and identify actions that might be taken to improve the safety of certain areas.
T3: Life Skills Module with Men and Boys	This started with the formation of groups of adolescent boys and young men (aged 15-25 years), who were recruited by the IP through community meetings and individual conversations. It aimed to increase men and boys understanding of underlying causes of VAWG and to build their capacity to challenge harmful social norms and take actions against VAWG at the community level. It included initial training for selected group members, a training for all members, regular monthly meetings and sporting events, and a range of community level activities and awareness raising events.

[16] While many of these groups were SHGs, in slums where these did not exist, kitty groups or women's micro-finance institution groups (MFIGs) were selected instead. For simplicity, these are referred to as 'SHGs' throughout the report. The strength of all three types of groups varied considerably, both in terms of how often they met and how effectively they were run. Therefore, in the early months of the intervention, significant efforts were made to bring all groups up to a minimum standard before core activities began.

[17] For ethical reasons, the SHG strengthening module also included information on VAWG referral networks, given the known high rates of IPV in the target areas and inconclusive evidence on the impact of women' economic empowerment interventions on levels of IPV.

The planned content of each of the above modules is outlined in more detail on in table 2 below.[18]

Table 2: Minimum package of core activities per intervention

Interventions	Structured activities
T1: SHG strengthening	1. 12 sessions on financial literacy at the slum level. 2. 3 days' training on book/ record keeping for 2 office holders per SHG 3. A 1-day workshop on banking for 1 member per SHG 4. Regular monthly/weekly SHG meetings for all members
T2: SHG Strengthening + VAW Module	1. 3 days' training of 2 community animators per SHG on 'Gender and VAW' 2. 2 days' training of 2 community animators per SHG on 'Addressing VAW' 3. Women's safety audits (WSA) in each slum 4. 5 day orientation workshop for all SHG members 5. 6 Thematic sessions on VAW 6. Up to 10 interface meetings with existing service providers to connect the SHGs and their members with services to prevent and respond to VAWG 7. 3 day helping skills workshops for selected SHG members 8. Exposure visits
T3: Life Skills Module with Men and Boys	1. 3 days' training for 1 youth ambassador per slum 2. A 1-day workshop for all members of each youth group[s?] 3. 10 community level thematic meetings of youth groups 4. Weekly sports activities at the community level 5. 6 inter-slum events

Programme beneficiaries

As outlined above, the Programme aimed to create spaces where groups of women and groups of boys and men could learn skills, increase their awareness and understanding of women's rights, the causes and consequences of VAWG, and support one another to take actions to prevent or respond to VAWG. The interventions also encouraged these groups to engage with others in their community, to encourage learning, reflection and dialogue. This essentially created **four distinct beneficiary groups** within the Programme:

- **Female direct beneficiaries** (FD) who were members of the SHGs;
- **Female indirect beneficiaries** (FI) who were members of the wider community within the slum;

[18] MPUIIP (2015)

- **Male direct beneficiaries** (MD) who were members of the boys'/men's groups;
- **Male indirect beneficiaries** (MI) who were members of the wider community within the slum.

Intended outcomes

During the inception phase for this evaluation, the evaluation team spent time with the IP to co-produce a **Theory of Change (ToC)**[19] for the Programme, identifying the main areas in which it was intended to effect change and pathways to change.

It was confirmed that the ultimate aim of the Programme was to reduce the prevalence of IPV and the prevalence of non-partner violence against women and girls in public spaces. Further, it was agreed that there were **four primary outcome areas** including the reduction in prevalence of:

- experience of physical and/or sexual IPV
- experience of violence and harassment in public spaces
- perpetration of physical and/or sexual IPV
- perpetration of violence and harassment against women and girls in public spaces.

In the initial formulation of the ToC a set of **intermediate outcomes** along the intended pathways to reducing VAWG were identified including individual attitudes and norms which condone VAWG; women's economic empowerment; community action against VAWG; and reporting VAWG. Following the baseline research, **intermediate outcome measures** were further refined and validated with experts in India[20], which the notable addition of social norms measures at endline.

Experience from VAWG programming and 'safe cities' work elsewhere on the challenges of shifting deeply entrenched harmful attitudes and behaviours suggested that monitoring these intermediate outcomes could help assess whether progress towards higher-level impacts, including a reduction in VAWG, could be expected and to provide vital insights into ***pathways to change***.[21] Given the delays in Programme implementation, which meant that in some areas the intervention lasted only 9 months, analysis of shifts in intermediate outcomes became increasingly important for the endline. Table 3 lists the final set of intermediate outcome measures[22], and the text below summarises the programme theory behind each component of the theory of change.

Table 3: Intermediate outcomes for the Safe City Initiative Impact Evaluation Endline

Components of the Theory of Change	Intermediate outcomes
1. Attitudes and social norms around VAWG	1.1 Gender equitable attitudes 1.2 Attitudes towards IPV 1.3 Attitudes towards public VAWG 1.4 Descriptive norms around IPV 1.5 Injunctive norms around IPV 1.6 Descriptive norms around public VAWG 1.7 Injunctive norms around public VAWG
2. Reporting of VAW	2.1 Women reporting IPV to police or protection officer

[19] See annex 2 for the Theory of Change of the SCI.

[20] Endline outcome measures were validated at a workshop with VAWG experts in New Delhi in August 2015. For a list of experts consulted and meeting agenda please see appendix 1.

[21] See, for example, Women in Cities International (2012) and DFID (2012).

[22] For details of outcome measures at endline, please see appendix 2.

	2.2 Women reporting public VAWG to police or protection officer 2.3 Descriptive norms around reporting IPV 2.4 Injunctive norms around reporting IPV 2.5 Descriptive norms around reporting public VAWG 2.6 Injunctive norms around reporting public VAWG
3. Women's economic empowerment	3.1 Women earning their own income 3.2 Women's control of household income 3.3 Women's role in household decision making
4. Mobility and feelings of safety	4.1 Women's mobility outside their slum 4.2 Women's feelings of safety in public spaces during the day 4.3 Women's feelings of safety in public spaces at night
5. Actions taken to address VAWG	5.1 Support from SHGs to women who experience IPV 5.2 Support from police to women who experience IPV 5.3 Support from SHGs to women who experience public VAW 5.4 Support from SHGs to women who experience public VAW 5.5 Actions taken to prevent or respond to IPV 5.6 Actions taken to prevent or respond to public VAWG 5.7 Recognition of the state as an actor to bring about social change 5.8 Willingness to engage others around the issue of VAWG

Individual attitudes and social norms around VAWG

The Programme's ToC is based on the assumption that a lack of action to prevent and respond to VAWG is a result of widespread tolerance of it[23] and gender inequitable norms and attitudes, which in turn are fuelled by a lack of awareness of girls' and women's rights, and a lack of understanding about what constitutes VAWG and information about what can be done to prevent it.

According to social norms theory, shifting individual attitudes towards VAWG is unlikely to be sufficient to address VAWG, if harmful social norms are prevalent. This includes both descriptive norms around VAWG – i.e. the idea that VAWG is typical, as well prescriptive norms – i.e. the belief that VAWG is appropriate.[24] The endline survey included innovative methods to measure both of these concepts, alongside measuring shifts in individual attitudes and behaviours.

The Programme aimed, through work with groups, to increase awareness and understanding that would reduce tolerance and promote positive action. Encouraging boys and men to challenge their own beliefs and behaviours was intended to lead them to not *want* to be violent; and raising

[23] See for example: Heise, L. (2011)

[24] MJ Alexander-Scott, E Bell, J Holden (2016)

awareness among communities was intended to reduce tolerance of violence and encourage action. So boys and men would also feel they could not 'get away' with violent behaviour, either because of reduced opportunities to be violent or because of the social consequences.

Reporting of VAW

The Programme aimed, through work with groups, to increase reporting of VAW, through raising awareness of local VAW response and services, and increasing social acceptance of reporting. The endline survey included questions to measure attitudes and norms towards reporting VAW in the future.

Women's economic empowerment

The Programme aimed, through women's SHGs, to economically empower women as a strategy to prevent VAWG. The underlying theory is that poverty and financial stress within the household increase the likelihood of violence in the home, and that women who lack economic independence have fewer options in terms of seeking help or leaving abusive relationships. However, existing evidence suggests that the relationship between women's income and their control of income on the one hand and the levels of violence they experience on the other are complex and multidirectional[25]. For example, women who are economically active outside the home are possibly at greater risk of experiencing violence and harassment in public spaces, and there is some evidence that they are also, at least in the short term, more likely to experience violence in the home if men perceive their own breadwinner role to be under threat (although in the longer term violence may decrease).[26] Through training and awareness raising activities with women, the Programme aimed to strengthen women's empowerment and autonomy and ability to positively negotiate new roles within the family, exemplified by increased control of their income and increased participation in household decision-making.

Mobility and feelings of safety

Boys' and men's perpetration of violence and harassment in public spaces not only affects those who directly experience it, it also has a more widespread impact by increasing women's fear for their safety, which in itself can negatively impact on their lives. This fear can lead to self-imposed restrictions on mobility. Moreover, other people's perceptions of women's safety, in particular husbands' perceptions, can also lead to the *imposition* of restrictions on women's actions and freedom of movement. This manifests itself not only in terms of reduced mobility among women, but also women needing to seek permission from others before they can leave the home, particularly when they are unaccompanied. Through awareness raising and discussion, the Programme was intended to challenge communities' perceptions of women's safety and, through community safety audits, to encourage reflection on alternative actions to address safety instead of restricting women's mobility.

Actions taken to address VAWG

Through increasing demand for services among communities and groups, alongside work improving the capacity of state institutions to respond to and address VAWG, the Programme aimed to increase community actions against VAWG. This included improving recognition of the state as an actor to bring about social change, and increasing individual actions against VAWG. Through community

[25] See for example: Vyas, S. and C. Watts. (2008).

[26] Brody C et al (2015)

actions the Programme also aimed to build momentum at the community level and increase people's willingness to engage others on the issue of VAWG.

3. Evaluation approach and methodology

This section provides information on the evaluation design and methodology, including ethical considerations, which have informed all aspects of the evaluation.

RCT factorial design

As outlined in section 1, the purpose of the evaluation was to seek evidence of the effects of the Programme and to attribute these to specific interventions or combinations of interventions. In order to enable this attribution, the evaluation is based on a factorial **RCT** with treatment effects estimated through the analysis of a mix of quantitative and qualitative data collected at baseline and again at endline. [27]

Randomisation provides significant benefits for assessing causal effects since it means there are no systematic differences between treatment and control areas, at least in expectation.[28] This enables changes in outcomes to be attributed to Programme interventions, as opposed to other factors. Given the focus of the SCI on working with women's SHGs and groups of boys and men at slum level, the slum was identified as the most logical unit of analysis. Slums were selected based on pre-existing geographically defined areas, which were already recognised politically and by community members themselves. The 250 slums involved in the MPUIIP – and therefore also in the SCI – were therefore randomly assigned to treatment and control groups (further explanation of randomisation is provided in appendix 3).

Since assignment to treatment was randomised, it is **possible to generate unbiased estimates of causal effects using endline data only**. The baseline data serves two functions: first to check balance in the assignment to treatment and control---as was assessed already in the baseline report--- and second to reduce variance in the estimation of treatment effects. Whenever possible the key estimates provided in the findings section of this report employ baseline measures of outcome variables as controls in an effort to minimize variance.

As described above, the evaluation does not simply seek to estimate differences in effects between a treatment group which receives the Programme and a control group which does not. Rather, **the evaluation is designed to assess which of the Programme's three interventions – or combinations of these interventions - are most/least effective**. A factorial design was therefore selected for the RCT. Factorial designs not only enable the effects of individual interventions to be identified, but also the effects of various combinations of these interventions. There are two types of intervention that relate to the SHGs: the SHG Strengthening Module and the VAW Module. In order to test the effectiveness of these interventions, two SHGs in each of the 250 slums received one of the following:

- The SHG Strengthening Module on its own (categorised as treatment type 'S')
- The SHG Strengthening Module plus the VAW Module (categorised as treatment type 'V')

[27] During the inception phase for this evaluation, many alternative designs for the RCT were examined and considered, each with different benefits and weaknesses. Some elements of alternative designs considered are listed in appendix 4, which summarises the main benefits they would have brought and the reasons why they were rejected in favour of the selected design.

[28] Imbalances can arise in practice which can produce "conditional bias." In the analyses such imbalances have been addressed on observables using regression techniques on pre-specified covariates.

- No intervention at all (categorised as treatment type 'N')

There is just one intervention which the boys' and men's groups could receive: the Life Skills Module. The men and boys group in each of the 250 slums could therefore receive one of following:

- The Life Skills Module (categorised as treatment type 'L')
- No intervention at all (categorised as treatment type 'N')

In order to enable comparisons of the various combinations of the three possibilities for SHGs and the two possibilities for the boys' and men's groups, a 3x2 factorial design was used, as outlined in table 4. The 3x2 factorial design produces six experimental arms. These are outlined in table 5.

Table 4: Factorial design for the evaluation of the Safe Cities Initiative

	Men and boys' groups			
Women's groups		**No Life Skills Module**	**T3: Life Skills Module**	**Total**
	No SHG intervention	41 slums 'Pure' control (Type 'NN')	41 slums Life Skills only (Type 'NL')	82 slums No SHG intervention (Type 'TC')
	T1: SHG Strengthening Module	42 slums SHG Strengthening only (Type 'SN')	42 slums SHG Strengthening + Life Skills (Type 'SL')	84 slums SHG Strengthening (Type 'TS')
	T2: SHG Strengthening Module + VAW Module	42 slums SHG Strengthening + VAW (Type 'VN')	42 slums SHG Strengthening + VAW + Life Skills (Type 'VL')	84 slums SHG Strengthening + VAW (Type 'TV')
	Total	125 slums No Life Skills (Type 'TN')	125 slums Life Skills (Type 'TL')	**250 slums** **(Total)**

Table 5: The six experimental arms under the factorial design

Treatment arm	Interventions	Type	Number of slums	Intervention(s) in each individual slum
T1 only	SHG Strengthening Module only	SN	42	Two SHGs will receive the SHG Strengthening Module but no intervention will be delivered to boys' and men's groups
T2 only	SHG Strengthening Module + VAW Module only	VN	42	Two SHGs will receive the SHG Strengthening Module and the Life Skills Module will be delivered to the boys' and men's groups
T3 only	Life Skills Module only	NL	41	No interventions will be delivered to the SHGs but the Life Skills Module will be delivered to the boys' and men's groups

T1+T3	SHG Strengthening Module + Life Skills Module only	SL	42	Two SHGs will receive the SHG Strengthening Module plus the VAW Module but no intervention will be delivered to the boys' and men's groups
T2+T3	SHG Strengthening Module + VAW Module + Life Skills Module	VL	42	Two SHGs will receive the SHG Strengthening Module plus the VAW Module and the Life Skills Module will be delivered to the boys' and men's groups
C	'Pure' control	NN	42	No interventions will be delivered
Total slums			250	

A key advantage of the factorial design is that data from each of the cells within table 5 can be used for multiple comparisons. For example, slums in the T3 group feature among treated slums when the effects of the Life Skills Module are examined, but feature among control slums when the effects of the VAW Module are examined. It is important to note that power is higher for comparisons of the individual interventions/ modules, rather than combinations. The set of priority comparisons are detailed in Table 6.

Table 6: Comparisons of interest within the factorial design

Comparison	Number of slums
T1	Data from 166 slums can be used to assess the effect of the SHG Strengthening Module, compared to no SHG intervention at all.
T2	Data from 166 slums can be used to assess the effect of the SHG Strengthening Module plus the VAW Module compared to no SHG intervention at all.
T3	Data from all 250 slums can be used to assess the overall effect of the boys'/men's Life Skills Module.

Evaluation hypotheses

Based on the three Programme interventions; the primary and intermediate outcomes to be achieved; and existing VAWG literature, a set of hypotheses was developed by the evaluation team in consultation with the IP, which were then tested through the RCT. These are outlined in table 7.

Table 7: Hypotheses to be tested through the RCT

Hypothesis identifier	Hypothesis
H1	Building the capacity of SHGs (T1) and their members leads to a reduction in reported experience of IPV and violence and harassment in public spaces by SHG members (direct beneficiaries) and by women who live in their slum (indirect beneficiaries)
H2	Building SHG members' understanding of – and ability to respond to - VAW (T2) leads to a reduction in reported experience of IPV and violence and harassment in public spaces by SHG members (direct beneficiaries) and by women who live in their slum (indirect beneficiaries)
H3	Building boys'/men's capacity to understand and advocate against VAW (treatment T3) leads to a reduction in reported perpetration of violence and harassment by members of the boys' and men's groups (direct beneficiaries) and by boys and men who live in their slum (indirect beneficiaries)

H4	Building boys'/men's capacity to understand and advocate against VAW (treatment T3) is *as effective* as building the capacity of SHGs and their members and building their understanding of – and ability to respond to - VAW (T1 and T2)
H5	The combination of these three interventions (T2+T3) leads to a reduction in reported perpetration of violence and harassment by members of the boys' and men's groups (direct beneficiaries) and by boys and men who live in their slum (indirect beneficiaries) and a reduction in reported experience of IPV and violence and harassment in public spaces by SHG members (direct beneficiaries) and by women who live in their slum (indirect beneficiaries)
H6	These three interventions (T1, T2, T3) are most effective when applied in combination (i.e. positive interactive effects)
H7	These three interventions (T1, T2, T3) lead to women SHG members (direct beneficiaries) and women who live in their slum (indirect beneficiaries) feeling safer in the home and/or in public spaces
H8	These interventions (T1, T2, T3) lead to SHG members (direct beneficiaries) and women who live in their slum (indirect beneficiaries) having greater mobility and use of public spaces, especially at night
H9	These interventions increase SHG members (direct beneficiaries) engagement with livelihood schemes
H10	These interventions increase SHG members' income levels and control over their income
H11	These interventions lead to increased autonomy among SHG members (direct beneficiaries) and women who live in their slum (indirect beneficiaries) in terms of freedom of movement and decision making in the home
H12	These interventions lead to a decrease in attitudes among SHG members and members of the boys' and men's groups (direct beneficiaries) and women and men who live in their slum (indirect beneficiaries) who blame women and girls for violence and harassment
H13	These interventions expand knowledge among SHG members and members of the boys' and men's groups (direct beneficiaries) and women and boys and men who live in their slum (indirect beneficiaries) of women's legal rights and the causes and consequences of VAWG
H14	These interventions lead to SHGs and their members and members of the boys' and men's groups (direct beneficiaries) and women and men in who live in their slum (indirect beneficiaries) becoming engaged in actions to prevent and respond to VAWG
H15	These interventions lead to increased reporting of VAWG and greater access to support for responding to IPV among SHG members (direct beneficiaries) and women who live in their slum (indirect beneficiaries)

Timeframe

Delivery of the SCI began in a limited number of slums from March 2014, with formal roll out after June 2014. Programme activities ended in November 2015.

Baseline quantitative survey data was collected over a four-month period between September and December 2013. Qualitative baseline data was then collected over eight weeks between January and March 2014. Endline quantitative and qualitative data was collected over a four-month period from November 2015-February 2016.

Mixed-methods approach

The evaluation employed **a mixed-method data collection strategy** involving a range of methods (as outlined in table 8) designed to provide comprehensive data to assess Programme impact, as well as provide useful insights around processes of change.

Table 8: Data collection methods at baseline and endline

Approach	Methods at baseline	Programme Monitoring	Endline
Quantitative	Panel Survey	Ongoing collection of Programme data from the SHGs and boys'/ men's groups using monitoring registries to capture activities	Panel Survey
			Behavioural measure
Qualitative	Focus group discussions including participatory techniques		Focus group discussions including participatory techniques
	Key informant interviews		Key informant interviews

Quantitative panel survey

The hypotheses, which have informed the factorial design, were tested using quantitative data collected through **a quantitative panel survey** with longitudinal data collected with sampled individuals at two points in time - baseline and endline.

The **baseline survey** instrument had been developed using a range of good practice from previous VAWG-related studies and evaluations, both internationally and specifically in India.[29] The survey instrument[30] was refined at endline taking into account the lessons learned at baseline and changes in the outcome areas of focus (including the addition of questions to measure shifts in social norms). It included questions focused on each of the primary and intermediate outcomes over which the Programme is intended to effect change. Almost all of the questions in the instrument were designed to capture self-reported experiences, attitudes, norms and behaviours. A number of questions were also included as proxies in order to triangulate and test the consistency of the data. In addition, the endline survey instrument included both an embedded list experiment and an innovative behavioural measure (described in more detail below).

The final survey instruments were translated into Hindi and programmed on Personal Digital Assistant (PDA) devices by NCIS.

List Experiment

In addition to the questions asking respondents to report their actual experience or perpetration of violence, the survey instrument included an **embedded list experiment**, intended to provide data on experiences and perpetration of violence without respondents having to directly answer questions about this. Such list experiments have been successfully used in previous studies to gather information on respondents' attitudes and behaviours which they do not readily admit to, for example racist or

[29] For example, the evaluation drew learning from the DFID What Works to Prevent VAWG Programme around common outcome areas, social norms theory, and internationally tested measures including the gender equitable men scale (GEM) developed by Promundo and the Population Council.

[30] See annex 3 for a copy of the complete endline survey instrument.

extremist views.[31] It involves providing respondents with a predefined list of behaviours and asks them to report *how many* behaviours from the list they have encountered, without specifying which ones.

Half the respondents receive a standard list of (usually) non-sensitive behaviours and it can then be calculated how many behaviours from that list are reported on average. The other half receives a longer list that includes both the standard list and the sensitive item (here, experience or perpetration of VAWG). The difference between the average number of affirmative responses provides an estimate of the proportion of respondents that have experienced or perpetrated the sensitive behaviour.

With regard to IPV, female respondents were asked about their experience of being slapped or having something thrown at them by a partner ("slapped"), and male respondents about their perpetration of the same behaviour in the last three months. In the case of VAWG in public spaces, female respondents were asked about their experience of being touched, groped, stalked or flashed at in a public place ("groped"), and male respondents about their experience of perpetrating the same. In the table below, we report the proportion of respondents in each beneficiary group who directly report experience/perpetration of these specific acts and the corresponding estimated proportions for the same group using the list experiment.[32]

Table 9: Estimated proportions of men and boys perpetrating VAW

Beneficiary	Slapped Partner (Direct Measure)	Slapped Partner(LE)	Groped in Public (Direct Measure)	Groped in Public (LE)
MD	0.065	0.2	0.03	0.426
MI	0.046	0.19	0.032	0.184

Table 10: Estimated proportions of women experiencing VAW

Beneficiary	Slapped by Partner (Direct Measure)	Slapped by Partner(LE)	Groped in Public (Direct Measure)	Groped in Public (LE)
FD	0.05	0.141	0.074	0.174
FI	0.07	0.105	0.081	0.237

There are clear differences in the estimated proportions of beneficiary groups who experienced or perpetrated the specific acts of VAWG from direct and list experiment measures, with proportions from list experiment being much higher. This suggests underreporting of sensitive behaviours in the direct measures. Both the direct measures and list experiment measures are used in the main analysis.

Note we describe a validation of the list experiment in Part C on measurement weaknesses.

Behavioural measure

At the end of the survey the enumeration teams implemented a module that generated a set of direct behavioural measures on **willingness to take action to address VAW** in their slums. The behavioural measure is motivated by concerns over reporting biases that can arise in surveys and seeks to find direct evidence of changes in behaviour. It focuses on a key interim outcome - willingness to seek change - which may be a more achievable goal than reducing actual violence for a programme of this form over a short time window.

[31] See for example: Corstange, D. (2009)

[32] Note that whenever we report or analyse the list experiment measure of perpetration of public violence by men, we restrict the data to surveys conducted after 18 Dec 2015. This is because there was a translation error in the original list experiment text, which was subsequently corrected on this date.

The measure was introduced by reminding subjects that the city's municipal corporation has the power to allocate their efforts and money in their budget across different issues. However, they have limited time and funds and have to make choices. Putting more money and more time towards one thing means putting less money and less time towards another. Subjects were then told what the three biggest areas of expenditure were for their city. They were told that some people feel that more funding needs to be directed towards addressing VAW (for half the subjects the focus was on IPV, for half on VAWG in public spaces). They were then given a card on which they could mark what position they thought the government should take on the use of funding. They could do this in relative privacy (the enumerator turned their backs as the subjects filled out the cards) before returning it to the enumerator who then placed the card in an envelope. They were advised that for this item their response would be delivered to their ward member so he/she knows what people in the slum think. In half the cases, subjects were asked to sign the card and in half not.

In addition, they were given an envelope with ten cards and invited to seek signatures from other slum residents (men and women) to indicate their support for increased spending to respond to VAW. The cards were colour coded to distinguish between male and female signatories, and respondents were advised not to gather signatures from individuals that had already signed a card.

The cards, whether completed or not, were collected at the end of the day or the following day by enumerators. The signed cards were then counted, providing a measure of both the share of men and women in a slum that are willing to take a public action against VAW, as well as their willingness to publically engage others on the issue.

In theory, all the treatment arms could have affected willingness to engage on this issue. T2 and T3 are meant to raise awareness around the issue, whereas T1 is meant to empower women in general and perhaps make them more active on issues they already care about.

Quantitative training and data collection

The quantitative fieldwork was managed by a Delhi-based team from NCIS. The team oversaw the selection of experienced enumerators and field supervisors for the survey, who were recruited within Madhya Pradesh. [33] A larger number of enumerators than needed was recruited and trained, enabling only those with the strongest skills to be employed for data collection.

Enumerators received a 10-day training course[34] delivered by NCIS together with members of the evaluation team, which comprised a mix of classroom work and practical exercises, some under real field conditions. It included a two-day sensitisation session on gender and VAWG led by Jagori, a Delhi-based women's rights organisation, and training on research ethics designed by the team at SDDirect.

All surveys were conducted using PDAs to reduce enumerator error. Throughout data collection, raw data was checked at regular intervals by the team at Columbia University with input and guidance provided on a regular basis. In addition, a quantitative research assistant – Sarah Khan - from Columbia University oversaw the training, pre-testing and data collection and was able to troubleshoot in real time with NCIS in the field.

Sampling for the survey

As discussed in section 1, four beneficiary groups were identified as most likely to benefit from the Programme. These four groups were identified prior to the baseline in all slums.

- Female direct beneficiaries (FDs) who were members of the SHGs;

[33] Many of the enumerators and supervisors at endline were also involved in the baseline data collection.

[34] Delivered in Bhopal & Gwalior: 26th Nov to 5th Dec 2015, and Indore & Jabalpur: 6th to 15th Dec 2015

- Female indirect beneficiaries (FIs) who were members of the wider community within the slum;
- Male direct beneficiaries (MD) who were members of the boys'/men's groups;
- Male indirect beneficiaries (MI) who were members of the wider community within the slum.

Including all four of these groups in the evaluation allowed for both the immediate effects of the Programme on direct beneficiaries where we might most expect to see change and as well as wider impact on indirect beneficiaries to be assessed.

The age ranges for the direct beneficiaries sampled for the evaluation were dictated by the age groups being targeted by the Programme. Choosing to work through existing SHGs meant the Programme would reach women over the age of 18 years. For the new boys'/ men's groups, the Programme deliberately targeted a younger age group: those aged 15 to 25 years. The sampling of direct male beneficiaries for the evaluation mirrored this, although in-line with other VAW-related research and evaluations in India and elsewhere, an upper age limit of 49 years was set for the other three beneficiary groups.

Based on the Programme's theory of change, it was expected that the indirect beneficiaries most likely to benefit from the SHG interventions (in particular the VAW Module) would be women and men of a similar age to the direct beneficiaries (i.e. 18 – 49 years). It was also expected that the indirect effects of the boys'/ men's Life Skills Module would be concentrated among boys and men of the same age (i.e. 15-25 years). For this reason, in each slum, men and boys mirroring the age group of direct beneficiaries (15-25) as well as older men and boys (25-49) were randomly sampled. However, for purposes of analysis, we pool these two age groups to maximise power.

The age groups selected for the direct and indirect beneficiary groups included in the evaluation are shown in table 11.

Table 11: Age ranges for direct and indirect beneficiaries

Beneficiary group	Age range
Women direct beneficiaries	18 – 49 years
Boys and men direct beneficiaries	15 – 25 years
Women indirect beneficiaries	18 – 49 years
Boys and men indirect beneficiaries	15 – 49 years

Two different sampling approaches for the survey were developed and used at baseline – with the aim of re-interviewing the same participants at endline: one for direct beneficiaries (male and female) and one for indirect beneficiaries (male and female). These are described below:

- Direct beneficiaries were identified from lists provided by GHK/IPE Global of five women from each of the 500 SHGs spread across the 250 slums, and on lists of eight boys and men who signed up to join a men and boys group from across the 250 slums. When more names than needed were provided, simple random sampling was used to select subjects.
- Selection of indirect beneficiaries was more complex, requiring individuals from three population groups to be identified: boys/young men (15-25) (a comparable age to male direct beneficiaries), older men (26-49) and women (18-49). This involved a two-stage process. In the first stage, research teams constructed a map of each slum in which every household was numbered consecutively. Using independent random number tables for each slum,

households were then selected for possible inclusion of an occupant from one particular population group. Enumerators then visited the selected households to determine whether there was someone from their target population they could interview. If there was not, the enumerators selected the next household number on the map and repeated the process. In the second stage, random sampling tables (unique for each household) were used to select individual respondents from a list of eligible household members.

- Overall, the endline survey had an attrition rate of **10.01**% i.e. **10.01**% of original baseline respondents could not be reached. Importantly, we find that attrition rates are balanced across treatment groups. Replacement respondents were selected every time an original respondent was unavailable. For more detail on the replacement strategy and implications for analysis please see appendix 5.

It is important to note that all survey respondents were sampled prior to randomisation and delivery of the programme. Therefore "direct beneficiaries" in slums that were not randomly selected to receive the relevant intervention (i.e. female direct beneficiaries in slums that did not receive either the SHG Strengthening or VAW treatment, and male beneficiaries in slums that did not receive the Life skills treatment) are better understood as "potential direct beneficiaries". As they were sampled in the same way from lists of SHG members and lists of men/boys who expressed interest in participating in the programme at baseline, these "potential direct beneficiaries" serve as a valid comparison group to the actual direct beneficiaries in slums that were randomly selected to receive the relevant intervention.

Table 12 outlines the intended sample of direct and indirect beneficiaries in each of the 250 slums, with an overall target of 7,500 respondents.

Table 12: Distribution of survey respondents per slum

	Boys/ young Men (15-25 years)	Older Men (26-49 years)	Women (18-49 years)	Total
Direct beneficiaries from SHG 1			4	4
Direct beneficiaries from SHG 2			4	4
Direct beneficiaries from the boys'/ men's group	6			6
Indirect beneficiaries from the wider slum population	4	4	8	16
Total per slum	10	4	16	30
Total for all 250 slums	2500	1000	4000	7500

An actual baseline and endline sample of 7,486 and 7,480 respondents respectively was achieved. These numbers correspond almost exactly to the target sample, as outline in table 13, and matched the planned distribution set out in table 12.

Table 13: Comparison of target and actual survey sample at baseline

Beneficiary group	Target sample for baseline	Actual baseline sample	Actual endline sample
Women direct beneficiaries	2,000	1,996	1,996
Women indirect beneficiaries	2,000	1,998	1,998

Boys/ men direct beneficiaries	1,500	1,500	1,498
Boys/men indirect beneficiaries	2,000	1,992	1,988
Total	7,500	7,486	7,480

Analysis of the endline survey data

The quantitative analyses reported in Part B estimate treatment effects using the following model:

$$Y_{i,t} = \beta_0 + \beta_1 T_1 + \beta_2 T_2 + \beta_3 T_3 + \beta_4 T_1 T_3 + \beta_5 T_2 T_3 + \delta Y_{i,t-1} + \gamma X + \varepsilon_{l,t}$$

Combination of coefficients from this model allow us to estimate a variety of treatment effects from the intervention including average (main) treatment effects, conditional effects, and interaction effects. For the purpose of analysis we refer to the three treatments as T1 (SHG only), T2 (SHG+VAW module) and T3 (Life skills (men and boys) module). The table below shows the mapping between nine treatment effects and regression coefficients. The most important of these are the average effects and the interaction effects.

Table 14: Mapping between causal effects and regression model coefficients

Causal Effect	Interpretation	Regression Coefficients
T1 Effects		
b(T1 | T3 0)	**Conditional effect:** The effect of SHG strengthening when there are no VAW or life skills (men and boys) interventions.	β_1
b(T1 | T3 1)	**Conditional effect:** The effect of SHG strengthening when there are no VAW interventions but there are life skills (men and boys) interventions.	$\beta_1+\beta_4$
b(T1)	**Average effect:** The average effect of SHG strengthening when there are no VAW interventions (over cases with and without men and boys interventions)	$\beta_1+.5\beta_4$
T2 Effects		
b(T2 | T3 0)	**Conditional effect:** The effect of SHG strengthening plus VAW interventions when there are no life skills (men and boys) interventions.	β_2
b(T2 | T3 1)	**Conditional effect:** The effect of SHG strengthening plus VAW interventions when there are life skills (men and boys) interventions.	$\beta_2+\beta_5$
b(T2)	**Average effect:** The effect of SHG strengthening plus VAW interventions averaged over cases in which there are and are not life skills (men and boys) interventions.	$\beta_2+.5\beta_5$
T3 Effects		
b(T3 | T1 & T2 = 0)	**Conditional effect:** The effect of life skills (men and boys) interventions only, when there are no interventions with women's groups.	β_3
b(T3 | T1 =1)	**Conditional effect:** The effect of life skills (men and boys) interventions in slums in which there is also SHG strengthening.	$\beta_3+\beta_4$

b(T3 | T2 = 1)	**Conditional effect:** The effect of life skills (men and boys) interventions in slums in which there is also SHG strengthening plus VAW interventions.	$\beta_3+\beta_5$
b(T3)	**Average effect:** The average effect of life skills (men and boys) interventions across cases in which there are and are not interventions with SHGs.	$\beta_3+\beta_4/3+\beta_5/3$
Interaction Effects		
b(T1T3)	**Interaction effect:** The increased effect of the life skills (men and boys) interventions when there is SHG strengthening compared to when there is not (and vice versa)	β_4
b(T2T3)	**Interaction effect:** The increased effect of the life skills (men and boys) interventions when there is SHG strengthening plus VAW interventions compared to when there is not (and vice versa)	β_5

The approach applied to model estimation is standard, including baseline data but no post-treatment data, a limited set of controls, with imputed data for controls when this is missing, and clustering standard errors at the level of treatment.

- **Baseline data**: Given the experimental design the difference in changes over time between treatment and control can be estimated by comparing endline measures in these two groups. In this context baseline data is used in two ways. First it was used in order to generate blocks prior to randomisation. This maximises slum level balance on these baseline measures. Second, baseline data are introduced as control variables on the right hand side of the regression. Introducing these measures on the right hand side provides a flexible way of using the baseline data. In those cases in which there is access to lags of the outcome variables this approach does not constrain coefficients on the lags to be one. In those cases where question wording has changed or new items have not been introduced and there is no access to lags, it is still possible to introduce related baseline data on the right hand side that can control for baseline variance.
- **Imputation:** Data for treatments or outcomes is not imputed, however in cases with missing data on control variables - which arises especially for baseline data for replacement subjects – data is imputed for the slum level – respondent-type mean value. In such cases a dummy variable is included that records whether there was imputation or not.
- **Controls:** As controls baseline levels of the outcome are included---or measures closely related to these---whenever these are available as well as enumerator fixed effects, pre-treatment predictors of VAW that were specified in advance in our analysis plan, block fixed effects, an indicator for whether an endline respondent replaced a baseline respondent, and an indicator for whether any covariate means were imputed for a respondent.
- **Standard Errors:** Robust standard errors are reported with clustering at the slum level. This takes account of the fact that the treatment was implemented at the slum level; moreover it does not impose a homoscedasticity assumption and returns the design-based "Neyman" standard errors[35] appropriate given this experimental design.
- **Analysis of list experiment.** The analysis of the list experiment questions for sensitive items is similar except that the coefficients of interest are interactions between treatment conditions and the "long list" treatment. See Box 2.

[35] Neyman standard errors are estimates of uncertainty that can be calculated using information from the randomisation procedure and do not require assumptions about the distribution of errors.

Box 2: List Experiment Analysis

Lists experiments provide an indirect way to measure a sensitive item. Half the respondents receive a standard list of (usually) non-sensitive behaviours and it can then be calculated how many behaviours from that list are reported on average. The other half receives a longer list that includes both the standard list and the sensitive item (here, experience or perpetration of VAWG). The difference between the average number of affirmative responses provides an estimate of the proportion of respondents that have experienced or perpetrated the sensitive behaviour. Thus the "longlist *effect*" is a measure of the *level* sensitive outcome. Then the effect of a treatment, *T*, on the behaviour is measured by assessing how *T* alters the longlist effect. This is given by the interaction between T and a long list indicator.

Say for example that in the control condition (T=0) the average answer to the list question was 2.5 for those getting the short list and 3 for those getting the long list, and in the treatment condition (T=1) the average answer was 2.5 for those getting the short list and 2.7 for those getting the long list. Then the estimate of VAW in the control condition would be 3 – 2.5 = .5 = 50%. The estimate in the treatment condition would be 3 – 2.7 = .3 = 30%. And so the estimated effect of T is 50% - 30% = 20%.

In a regression framework we estimate:

$$\text{Number answered} = a + b\ \text{LongList} + c\ \text{T} + d\ \text{LongList} * \text{T}$$

Coefficient *b* is then the level of VAW in control, *c*+*d* is the level in treatment, and *d* is the effect of treatment.

Qualitative component

In addition to the quantitative survey, the evaluation included a qualitative component to support interpretation of survey data and, critically, to help understand how and why change may or may not have occurred. Consequently, the qualitative component of the evaluation had the following **three objectives**:

- **To generate qualitative data to supplement, explain and triangulate with quantitative endline data**. The qualitative data is intended to help explain the quantitative findings – to develop a much deeper understanding of how the intervention(s) were received, and barriers/enablers to change.
- **To develop a more comprehensive picture of the Programme's performance in line with its ToC across a wider range of outputs, outcomes and impacts**. Qualitative methods were used to generate detailed insights into processes of, motivations for and barriers to change. They also helped track how the intervention(s) were working, helped understand the relative importance of intervention techniques or components, and helped answer "how" and "why" questions of the evaluation.
- To make a broad assessment of the **effectiveness, relevance, value for money (VfM)[36],** and **sustainability** of the programme.[37]

[36] The VfM analysis is provided in annex 6.

[37] OECD-DAC Criteria.

A summary of the qualitative measures used at endline is provided in appendix 5 and analysis of the qualitative data is integrated within Part B. An example FGD guide and KII guide are included in annexes 4 and 5.

FGD sample

The endline sample included 48 focus group discussions (FGDs)- 12 per city, divided equally across the four cities (see table 16). Slums were selected using purposive sampling to include a cross-section of direct beneficiaries across five of the six treatment arms: T1 (SHG only), T2 (SHG+VAW), T3 (life skills/men and boys), T2+T3, and control.

Table 15: Endline FGD sample in each city

Treatment arm	SHG only		SHG + VAWG only		Men and boys only		All 3 interventions		Control
Slum	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Direct female beneficiary	1 FGD	1 FGD	1 FGD	1 FGD			1 FGD	1 FGD	1FGD
Direct male beneficiary					1 FGD	1 FGD	1 FGD	1 FGD	1 FGD
Total FGDs per city	12 FGDs								

The sampling strategy for the qualitative component was designed to explore any differences between direct beneficiaries who received the three core modules individually in comparison to control. We also included all intervention slums in order to broadly look at interactions between the interventions.

The focus was on direct beneficiaries with the rationale that those who were direct beneficiaries of interventions is where we would expect to see most change as well as to be able to explore beneficiary perceptions of implementation and delivery.

We purposively selected FGD participants from high performing SHGs/Men's and boys' groups (based on monitoring data and discussions with GHK/IPE) in order to enable an assessment of the programme's ToC and underlying assumptions. Each focus group included 8-10 participants.[38]

Key informant interviews

Nineteen KIIs were conducted at endline with key stakeholders including representatives of implementing partners, SCI Programme staff, NGOs, local government, police and wider stakeholders across each city. They captured a wide range of views, including from individuals who played a direct role in overseeing or implementing the Programme, managers and providers of key services for women and girls, and other individuals with knowledge of and/ or influence over the situation of women and girls in the slums. For a list of the KIIs conducted, see appendix 7.

[38] In the treatment slums, women's FGDs included participants who were members in either of the two functioning SHGs from each slum. In control slums, women participants were also members of SHGs, though not part of the intervention.

Qualitative training and data collection

The qualitative data collection was led by a small team of experienced qualitative researchers from Jagori. FGDs were undertaken by teams of two researchers (the same sex as respondents), including a note taker and an experienced facilitator. KIIs were undertaken by one researcher. All researchers undertook a four-day training including a practical field test.

FGD were developed in line with the outcome areas of interest. The FGD guides included a range of participatory exercises and games, designed to elicit rich discussions around the key outcome areas of focus. These included video vignettes to stimulate discussion around VAWG, visual circles of change diagrams to assess individual perceptions of most significant changes as a result of the programme, and a participatory bead exercise to assess perceptions of levels of VAWG.[39] FGDs were also recorded using digital recorders with permission from participants. All qualitative instruments were translated into Hindi by Jagori.

The pre-testing of the instruments was conducted over four days in October 2015, and the qualitative discussion guides and translations further refined. Following data collection, the FGD recordings were transcribed and translated into English by a team of transcribers and translators at Jagori.

Analysis of the FGDs and KIIs

The main approach to the qualitative data analysis at endline was deductive: grouping and analysing data in line with the outcome areas of interest to the evaluation. However, an inductive approach was also used to look for, and explore, other emergent issues and relationships. The qualitative data analysis included the following steps:

Step 1: Organisation and initial review of the data. An initial broad coding framework was developed based on the primary and intermediate outcome areas of interest to the evaluation as well as some open codes for other (unanticipated) themes. A sample of transcripts was used to refine these codes, including by sub-dividing key themes into sub categories.

Step 2: Coding of data. Once the coding framework had been finalised, the 48 FGD transcripts and 19 KII transcripts were coded by the SDDirect team – using a double approach of manual coding and inputting coded data into an Excel database. Regular dialogue between those conducting the analysis was conducted to ensure consistency.

Step 3: Synthesis of findings. Once the coding was complete, descriptive analysis was undertaken based on the outcome areas of interest. This involved extracting relevant data and citations to draw out key findings and to look for areas of agreement/disagreement. Significant extracts of the text were coded each time so that when it was analysed, there was adequate context to judge participant responses in light of the facilitator's question and prior discussion. On this basis, decisions were taken about whether to use the data for the analysis without reservation or to use it in a more limited way (e.g. where a leading question had been asked). As is common practice with qualitative data analysis, these assessments of quality were used to make judgements throughout the analysis process about the meaning, validity and consistency of data from different FGDs and KIIs.

Step 4: Interpretive and comparative analysis. In-depth interpretive analysis was then conducted, drawing out relationships between different outcomes and explaining and contextualising these with reference to wider theory, research and evidence on VAWG. This stage included content analysis of transcripts to explore gender and social norms. With regard

[39] See annex 4 for example FGD guide.

to the latter, particular attention was paid to normative and empirical expectations of others, evidence of positive sanctions for compliance and evidence of negative sanctions for non-compliance.[40] Comparative analysis was used to assess differences between the intervention types. Findings were carefully cross-checked, which included sharing the analysis with members of the field research team for feedback and validation.

Step 5: Integration of qualitative and quantitative analysis. A systematic comparison of the findings of the qualitative and quantitative data was then conducted, guided by comparison of whether the qualitative data:

- Corroborated or contradicted the quantitative findings **(triangulation)**
- Helped to explain a pattern or correlation in the quantitative data – i.e., how and why change occurs more in one intervention than another **(explanation)**
- Pointed to new issues, patterns, unintended consequences, cause-effect relationships, impacts and explanations that were not identified by the quantitative analysis **(expansion).**

Ethical considerations

Ethical considerations are paramount when researching VAWG[41], and ethics were therefore of the utmost importance in this evaluation.

As part of the design phase for the evaluation, two international ethics experts were commissioned to review the proposed methodology from an ethical viewpoint and make recommendations to strengthen it. The approach sought to minimise reporting bias whilst ensuring the safety of field researchers and ensuring research participants were not put at increased risk. The approach was also mindful to not be exploitative of participants, and that access to support services was facilitated. A detailed table outlining the ethical guidelines for the evaluation are included in appendix 8. The core aspects of the approach to ensuring ethics were upheld at baseline and endline and are summarised below:

- A**ll field researchers received training on ethics and child protection and on women's rights and VAWG.**
- All interviews and FGDs were conducted/ facilitated by r**esearchers who were of the same sex as the participants.**
- **Field researchers were trained to terminate or change the subject of discussion if the interview was interrupted by anyone.** Privacy during interviews and group discussions was particularly difficult to achieve in a heavily built up urban environment. Efforts were made to arrange interviews and group discussions at times that suited participants. However, it often proved difficult to achieve full privacy given that most interviews were conducted in people's own homes, which often lacked space for a confidential conversation.
- **No more than one member of a household was interviewed** so that other household members would not be aware of the exact content of the survey. Boys/ men who were interviewed for the evaluation were also asked questions about their experiences, mobility and feelings of safety in their slums so that the survey was not obviously specifically focused on VAW.
- **Informed verbal consent was requested from all research participants**. It was not considered appropriate to ask for written consent given likely illiteracy among participants[42] and the concern

[40] Framework for measurement of Social Norms adapted from Mackie et al (2015)

[41] DFID (2012)

[42] Illiteracy among women is as high as 30% in the cities from which the sample was drawn. Total literacy rate in Bhopal is 82% (male 87.% and female 77%), Gwalior is 87.% (male 91% and female 79%) Indore is 87% (male 92% and Female 83%), Jabalpur is 75% (male 79% and female 70%) Indian Ministry of Home Affairs Census Bureau. (2011). 'Census of India'.

that getting signed consent could deter them from opening up about sensitive subjects. Verbal consent was therefore sought. Consent from parents or caregivers for the participation of those aged 15-17 years was not sought, and the emphasis was instead placed on ensuring that participants had sufficient information to provide consent themselves. A copy of the consent statement read out to participants was not left with them in case it was read by others.

- **Researchers were trained to detect signs of distress or trauma** and to pause or stop interviews or discussions and provide information on support services when necessary.
- **No names, addresses or other details that could allow identification of participants were recorded in the completed survey or in FGD notes and transcripts.**
- **All participants were given an information card with numbers of local support services**, including those responding to VAW. This card included a range of other services, so it would not arouse suspicion if seen by another family or community member. Researchers pointed out one main VAW service provider on the list to respondents who were illiterate.

Limitations

The previous section presented the evaluation approach, with an emphasis on the strengths of the factorial design and selected methods in terms of enabling an accurate estimation of treatment effects and exploring mechanisms and pathways to change. However, it is also important to highlight a number of caveats to the evaluation and limitations in terms of the selected approach. These are outlined in the following section. Some of these relate specifically to the evaluation of the SCI, while others reflect the limitations of VAW evaluations more widely.

Limitations of reliance on self-reported attitudes and behaviour

It is difficult to get accurate data on VAW, partly because it tends to rely on self-*reported* violence and partly because this reporting relies on people speaking openly about a very sensitive topic. The main data collection method for the RCT, namely the individual panel survey, relies on women telling enumerators about their experiences of violence and men telling them about the violence they have perpetrated.

Relying solely on the survey measure of self-reported experience (for female respondents) and perpetration (for male respondents) of VAW as an outcome measure for assessing Programme effects has the following limitations:

- The programme interventions encourage women to recognise certain behaviours as violence, to speak out about VAW and to report violence to relevant authorities. Since the Programme encourages this discussion, it may be expected that even if the programme achieves its goal of reducing the incidence of VAW in treatment communities it will simultaneously increase the reporting of VAW. Thus the survey might find higher reported levels of VAW in treatment communities than control communities due to greater awareness and willingness to report as a result of the treatment. The risk for this is especially high since on average only 12 per cent of respondents reported experience of VAW in the baseline survey[43], which is suggestive of a general trend of underreporting that may change as a result of treatment.
- A large part of this intervention focuses on attitudinal change and norm shifts in the community. However, self-reported survey measures of attitudes are subject to various types of response bias. For example, individuals may respond in a way they think the enumerator wants them to. An anonymous survey can also only capture the attitudes that people are

[43] Neville S, Mclean Hilker L, Humphreys M, Husain S, Khan S, Lindsey S (2014)

willing to express in private, while an important outcome of interest (and both a prerequisite and indicator of social norm change) is people's willingness to speak out against VAW in public.

For these reasons an important addition at endline is the inclusion of a behavioral observation measure which relies less on direct reporting, as well as an 'embedded list experiment' within the survey instrument which was used to estimate the extent of underreporting.[44] Furthermore, the quantitative data is complemented by the qualitative component. Whilst the qualitative component also relied on women and men (separately) to discuss sensitive issues within a group setting, it provided us with rich data on social norms around VAWG. Whilst attitudes expressed in a group setting may not be reflective of individual attitudes, they can tell us a lot about what is typical and appropriate behaviour within reference groups (social norms).[45]

Spill-over effects are likely and need to be understood

The RCT has been designed to capture certain spill-over effects which are of particular interest, where, through direct beneficiaries, the programme has an effect on men and women in the wider slum community (defined as the indirect beneficiaries). The rationale for focusing on indirect beneficiaries within each slum's boundaries was described in section 2 and 3, and for a number of reasons 'the slum' is the logical unit of analysis for this evaluation. However, given the fact that many of the 250 slums are in heavily built up areas which are geographically very close to each other, the effects of the Programme in treatment slums could spill over into neighbouring slums. Individual direct beneficiaries might also have a network of family, friend or workplace contacts that extends even further.

This is more likely to affect the measures of public VAWG than the IPV measures. The baseline analysis found that almost three quarters of reported incidents of violence and harassment in public spaces had taken place outside slum boundaries.[46] This indicates that boys and men may already be travelling to neighbouring slums in order to perpetrate violence and harassment. Therefore the Programme interventions themselves may encourage this to happen, for example if it became more difficult for boys and men to perpetrate violence in their own slum, rather than being prevented, the violence and harassment could simply be displaced. This could mean that Programme effects may not be discernible: if boys and men reduce the violence and harassment they are committing in other slums, Programme effects will be underestimated. Conversely, if violence and harassment is displaced away from home slums, Programme benefits may be over-estimated.

In order to address this, direct enquiries to survivors of non-partner VAWG in public spaces about whether perpetrators from the same slum were included in the endline survey. Furthermore, the qualitative component also investigated these effects and provided deeper understanding of how, why and where changes may or may not have occurred.

Limited monitoring data

The evaluation team was unable to access original monitoring files and so several of our conclusions around explanations for the results could not be verified.[47] The lack of available monitoring data

[44] See Part A, section 3.5 for further reference to the embedded list experiment.

[45] Framework for measurement of Social Norms adapted from Mackie et al (2015) When analysing FGD transcripts phrases showing group agreement (beliefs about others); acceptance (normative expectations of others are legitimate) such as 'everyone agrees' (enough others in the reference group believe) are indicative of social norms.

[46] Neville S, Mclean Hilker L, Humphreys M, Husain S, Khan S, Lindsey S (2014)

[47] The SCI programme ended in December 2015, with all staff contracts terminated or staff moved to different projects, the project office was closed and all project documentation was handed over to the ULBs. Inevitably, during endline analysis in April/June 2015 questions arose that such data and information from project teams might have helped to answer. However,

restricted the evaluation team's ability to track implementation fidelity or verify concerns of implementation. Furthermore, the evaluation analysis would have benefited from more qualitative data including beneficiary feedback, observations of the quality of delivery, and monitoring of signs of backlash.

PART B: ENDLINE RESULTS

4. Sociodemographics

Socio-demographic information was gathered in the baseline survey, and select socio-demographic characteristics (such as age, marital status and religion) were collected again at endline. We report on baseline demographics since our sample was defined prior to treatment; moreover since these measures are gathered pre-treatment they are not themselves a function of treatment and can represent the treated population accurately in a way that endline measures cannot. The following description of respondents draws from the analyses and text in the baseline report.[48]

Differences between direct and indirect beneficiaries

The baseline data contained information on 7,486 individuals of which 47% are defined as direct beneficiaries. The definition of individual respondents as 'direct beneficiaries' was based on lists compiled by the IP for the purpose of sampling for the survey.

Knowing how the direct and indirect beneficiary samples differ is important for assessing the broader relevance of the evaluation findings at endline. While the indirect beneficiaries were randomly sampled from the general slum population, the direct beneficiaries were either women who were members of existing SHGs, or young men and adolescent boys who were approached by the IP to take part in the Programme and who flagged an interest in doing so. Ultimately there will be interest in knowing whether these interventions would be effective if delivered to the general population rather than the specific beneficiary pool selected here. The analysis therefore benefits from understanding whether and how the beneficiaries differed from the general population.

Table 16 provides information on the differences between direct and indirect beneficiaries at baseline. It illustrates whether direct beneficiaries were more or less likely to have certain characteristics (indicated by the absence or presence of a '–' sign, with presence of the sign indicating that they were *less* likely).

requests for such data after December 2015 in large part could not be met and most staff were not available for further consultations.

[48] Neville S, Mclean Hilker L, Humphreys M, Husain S, Khan S, Lindsey S (2014)

Table 16: Bivariate correlations of variables with 'direct beneficiary' status [49]

Correlates	Married /cohabiting Females	Unmarried Females	All Females	Male Youths (age 15-25)
Scheduled Caste/Tribe/OBC	-0.047	0.057	-0.032	0.054
Main Religious Group	0.003	0.061	0.015	-0.012
Household Wealth	0.009	0.007		
Disability	0.063	0.03	0.05	-0.065
Childhood Exposure to Violence	-0.004	-0.019		0.001
From MP	0.003	0.251***	0.017	0.131*
Level of Education	-0.002	-0.014	-0.006	0.007
Age in Years	0.012***	0.012***	0.013***	-0.002
Working for Income	0.207***		0.202***	-0.009
Number of Unions	0.03	0.238***		
Pregnant	-0.134**			
Number of Children	0.043***			
Ever Given Birth to a Son	0.183***			
Age at Marriage	-0.012***			
Dowry Paid and Not Satisfied	-0.054*			
Dowry Paid and Satisfied	0.005			
Spouse's Education Level	-0.007			
Spouse's Alcohol Use	0.035			
Nuclear Family	0.067***			
Spouse Work Status	-0.013			
Spousal Age Difference	0.004			
Access to Sanitation			0.01	0.027*
Married or Cohabitating			0.104***	-0.057
Slum Alcohol Use			-0.001	0.028
Slum Pct. Below Poverty Line			-0.003	-0.004
Slum Fractionalisation			0.002	0.005

[49] Refer to appendix 10 for an explanation of how these correlates are coded. The variable 'direct' is coded 0 if a respondent was surveyed as an indirect beneficiary and 1 if surveyed as a direct beneficiary.

Slum Number of Households			0	0.001
Slum Male Unemployment			0	-0.015
Slum Attitudes on VAW			0	0.001
Attitudes on VAW				-0.004
Alcohol Use				-0.094*
Indicators for statistical significance: * for 0.1, ** for 0.05, and *** for 0.01				

Table 16 suggests that FDs are more likely to be older, married/cohabiting, to live in a nuclear family and have more children than FIs. They are also more likely to work for income, but are marginally less likely to be educated. By contrast, MDs tended to be younger and are slightly more likely to have higher levels of education than MIs (although this is not significant). They are also slightly less likely to report drinking alcohol on a regular basis.

The following two graphs show the age distribution of survey respondents by beneficiary status based on baseline data.

Figure 1 depicts a spread of FDs and FIs across the age range of 18-49 years, but with a higher proportion of directs in upper cohorts and a higher proportion of indirect beneficiaries in lower cohorts. The mean age of FDs is therefore higher than that of FIs. Figure 2 shows the distribution of MDs and MIs. Reflecting the differing sample selection criteria, the vast majority of MDs are aged 15-25 years, whereas MIs are spread across a broader age range of 15-49 years.

Figure 1: Age distribution of female beneficiaries

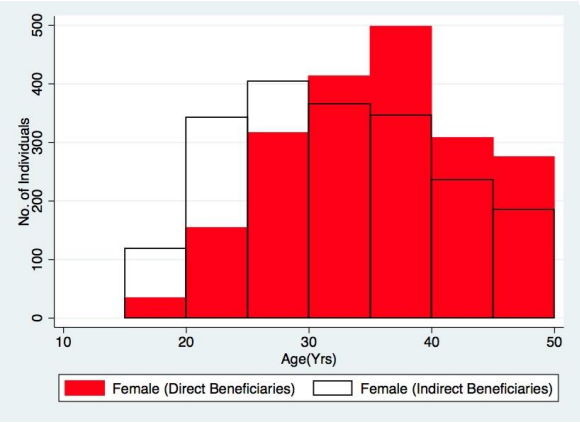
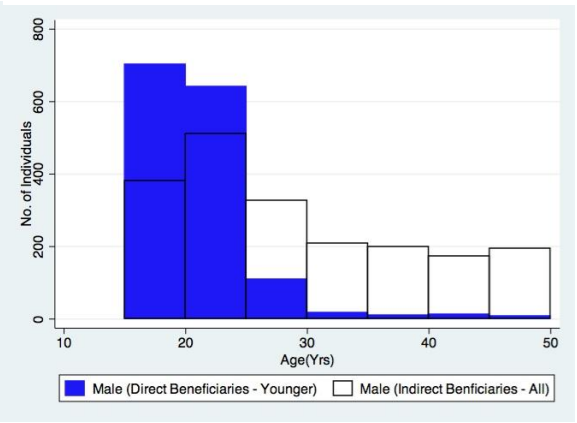


Figure 2: Age distribution of male beneficiaries



The following tables provide further comparisons between the direct and indirect beneficiary groups relating to religion and caste from baseline data.

Table 17 shows the distribution of religious affiliation among survey respondents at baseline. The vast majority of those surveyed (86%) are Hindu, with the remainder mostly Muslim. There was very little difference across the beneficiary groups.

Table 17: Religious affiliation among direct and indirect beneficiaries (%)

	Female Direct Beneficiaries 18-49 years	Female Indirect Beneficiaries 18-49 years	Male Direct Beneficiaries 15-25 years	Male Indirect Beneficiaries 15-25 years	Male Indirect Beneficiaries 26-49 years	Total
Hindu	87.3	86.6	86.3	86.8	86.7	86.8
Muslim	12.0	13.0	12.7	12.4	12.7	12.6
Buddhist	0.3	0.2	0.1	0.1	0.3	0.2
Sikh	0.4	0.1	0.4	0.2	0.2	0.2
Christian	0.0	0.1	0.2	0.2	0.1	0.1
Jain	0.0	0.1	0.3	0.1	0.0	0.1
No Religion	0.0	0.0	0.0	0.1	0.0	0.0
Other	0.1	0.0	0.1	0.0	0.0	0.0
Refused	0.0	0.0	0.1	0.0	0.0	0.0
Total	100	100	100	100	100	100

Table 18 shows the distribution of respondents according to caste. The overwhelming majority of respondents (83%) report to belong to a scheduled caste (SC), scheduled tribe (ST) or Other Backward Caste (OBC). These figures are fairly well balanced across the beneficiary groups.

Table 18: Caste affiliation among direct and indirect beneficiaries (%)

	FDs 18-49 years	FIs 18-49 years	MDs 15-25 years	MIs 15-25 years	MIs 26-49 years	Total
	Women (Direct beneficiaries)	Women (Indirect beneficiaries)	Young Men (Direct beneficiaries)	Young Men (Indirect beneficiaries)	Older Men (Indirect beneficiaries)	Total
Scheduled Caste	27.2	25.6	29.7	26.7	30	27.6
Scheduled Tribe	9.2	12.1	11.6	12.2	10.8	11.1
OBC	46.1	46.6	42.8	42	41.2	44.4
None	15.6	13.7	14.5	17.6	16.6	15.3
Refused	0.6	0.5	0.1	0.2	0	0.3
Don't Know	1.3	1.5	1.3	1.3	1.4	1.4
Total	100	100	100	100	100	100

5. Implementation checks

Programme take-up

At the time of the baseline survey, the pool of direct beneficiary respondents was randomly drawn from a list of potential beneficiaries who had been identified by the IP. They included women who were already members of existing SHGs or who had expressed interest in membership, as well as adolescent youth who had expressed interest in membership of boys' groups.

A consideration at baseline was the likelihood that direct beneficiaries would actually go on to participate in the Programme. The IP had asked all direct beneficiaries, prior to the random assignment of slums into treatment arms, whether they would be interested and willing to participate in the Programme and in the evaluation. This was asked of direct beneficiary respondents again in the baseline survey. Table 19 outlines responses in terms of how likely respondents said it was that they would actually take part in the Programme if it was delivered in their slum.

Table 19: Likelihood of participation in the Safe Cities Initiative

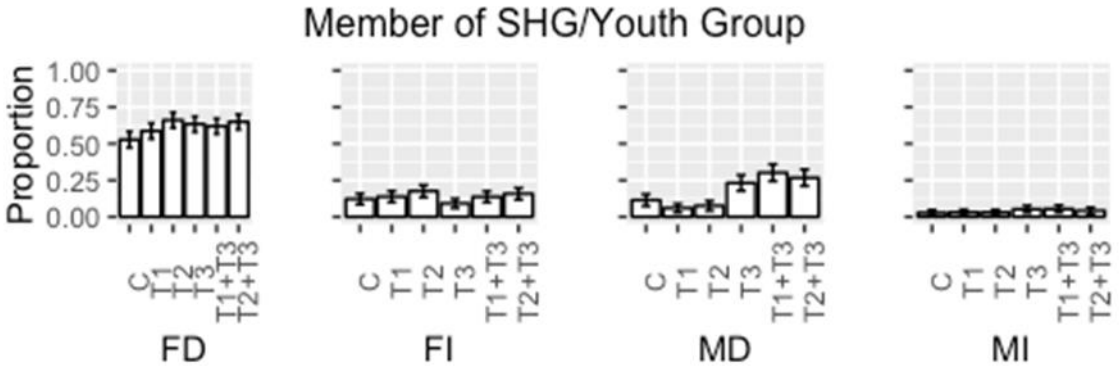
Likelihood of participating	Male direct beneficiaries	Female direct beneficiaries
Very likely	60%	60%
Somewhat likely	21%	30%
Not likely	18%	8%
Refused/ don't know	1%	2%

These figures show that the vast majority of direct beneficiaries said they were at least 'somewhat likely' to participate. However, they also suggest that the IP would need to invest in further engagement with some who said it was unlikely that they would take part. This applies to a larger proportion of boys and men than women.

At endline we asked those direct beneficiary respondents whether they are currently members of the SHGs and youth groups that were strengthened or formed anew as part of the Programme. We would expect to see higher membership rates in slums receiving a treatment.

The proportion of beneficiaries who report being a member of an SHG or youth group in their slum in each treatment arm is shown in the figure below.

Figure 3: Proportion of beneficiaries who report being a member of an SHG or youth group



Reported group membership for FDs is overall much higher than for MDs. A considerable proportion of FDs in the control group report being SHG members. This is not surprising and reflects the sampling strategy for FDs who were selected from members of SHGs, whereas boys and men's groups were created anew in all slums receiving the life skills treatment. As expected, membership levels are overall low (though non-zero) among indirect beneficiaries, male and female.

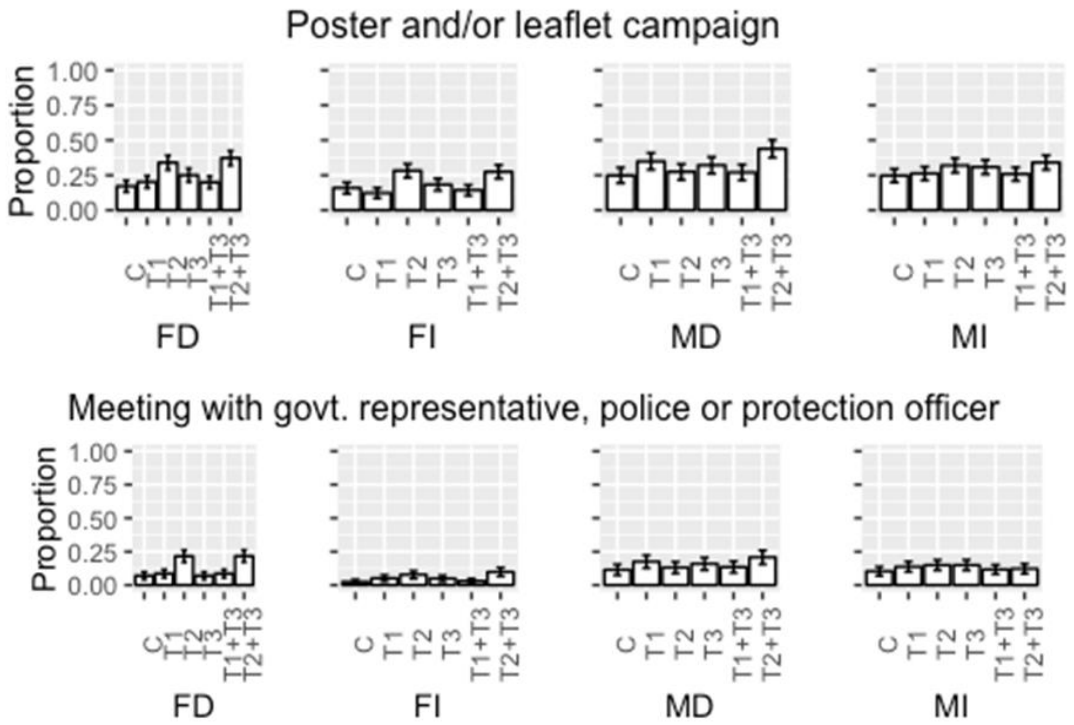
The proportion of MDs reporting youth group membership across all slums that had the life skills treatment (either in isolation or in combination with T1 or T2) is very low at **26.7** %. This may be due to low Programme take up among boys and men.

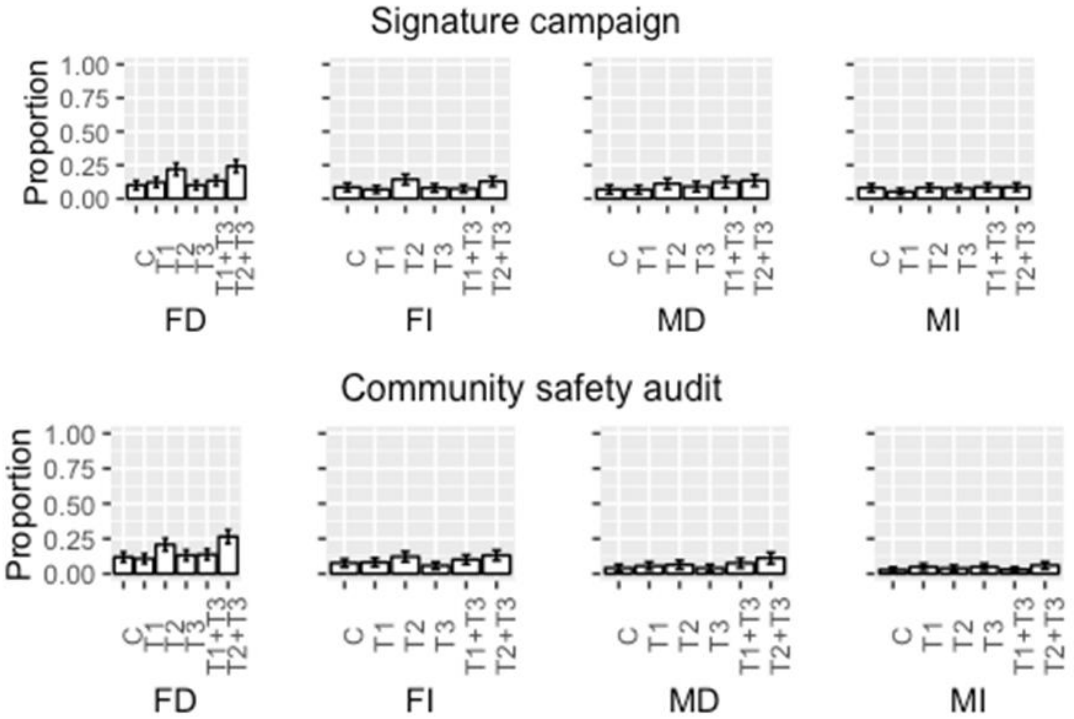
Indeed, if the interventions are intended to have impacts on direct beneficiaries and a considerable proportion of the direct beneficiary sample is in fact not a member of the group through which the intervention was delivered, it impacts our ability to identify Programme effects. To address this issue, we report results from a heterogeneous effects analysis by group membership.

However, at the same time, it is important to note that demonstrated inability to create and sustain wide group membership among the pool of direct beneficiaries speaks to a gap in implementation (discussed further in the conclusions).

Implementation checks

To assess implementation of the Programme at endline, respondents were asked of their knowledge of the SCI, as well as whether they were aware of any Programme-related activities that took place in their slum in the past year. The figures below present plots by beneficiary type and treatment arm for a selection of these implementation checks; the y axis denotes the proportion of respondents who were aware of such an activity taking place in their slum in the past year.





Programme recognition rates for the Safe Cities Initiative are under 50% for all populations, except among MDs in slums where the life skills treatment in isolation, or in combination with the SHG treatment, was applied. Recognition rates among respondents in control slums are not always significantly different from recognition rates in treatment areas. The reason for this may be that the initiative was publicised at the city level with banners and posters placed in public places across the city, not restricted to treatment slums.

Recognition of specific programme activities among beneficiaries is often very low and for no activity are recognition rates in excess of 50%. For example, the recognition rate for the women's safety audits, and meetings with government representatives and protection officers – activities which were cornerstones of the SHG strengthening intervention (T1) - are merely **23.6% and 21.7%** among FDs in slums that received the SHG+VAW module (T2), in isolation or in combination with T3.[50] These important findings will be considered later in the conclusions.

[50] It is possible that the translated term for WSAs was not well understood by survey participants and that enumerators did not sufficiently explain the concept. However, this finding is consistent with overall low rates of recognition across other activities.

6. Endline results by outcome area

Box 3: Guidance on reading the endline results

Part B, section 6 presents the main findings of the evaluation with respect to both the intended and unintended impacts and estimated effects associated with each intervention. These are organised around outcome areas, beginning with the primary outcomes and then moving on to intermediate outcomes. Under these main headings, related outcome measures have been clustered and are considered together to avoid repetition and to reflect their connectivity.

For each cluster of outcomes, the following information is presented:

- a description of the outcome measures included in the cluster
- for each outcome measure:
    - a narrative summarising the survey findings
    - results tables providing details of estimated effects of each treatment on all subgroups
- presentation and discussion of findings from the qualitative research.

An annotated example of results tables 'results at a glance' is provided below which explains how to interpret these tables, and a more detailed description is provided in appendix 10.

We present heterogeneous effects across cities in section 7.

Figure 4: Results at a glance

Description of the measure
Introduces the variable, how it was measured and what values the measure can take. The introductory text also links the measure to the survey instrument so that exact wording can be assessed.

Treatment arms
T1: SHG strengthening
T2: SHG+VAW
T3: Life skills (men and boys)
T1+T3: SHG strengthening + Life skills
T2+T3: SHG+VAW+ Life skills

Main Effects

Treatment Effect: Estimated average effects for each treatment (amount of change to the control mean value)

SE: Standard error of effects estimate.

Control Mean: Average estimate across all slums and treatments where the T is not implemented.

N: Sample size

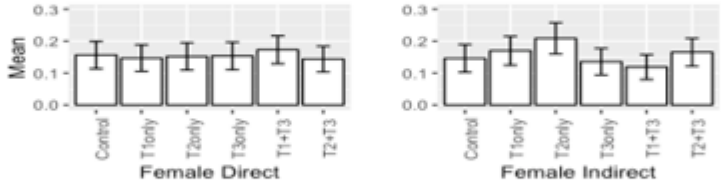
LongList: Estimated average outcome in the control group (list experiment

Treatment*Long: Effect of treatment (list experiment)

Interaction*long: interaction effects (list experiment)

Result 1: Women's experience of any form of physical or sexual IPV

Description of measure: This measures the prevalence of physical or sexual IPV among married or cohabiting women (EIPV_ANY). It is coded as 1 if the respondent has experienced at least 1 of 7 forms of physical or sexual IPV in the last 3 months using EDV7-EDV13 in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	-0.006	-0.017	0.002	Interaction Effect	0.054	0.018
	SE	0.022	0.021	0.017	SE	0.042	0.042
	Control Mean	0.151	0.157	0.152	Control Mean	0.156	0.156
					T1 or T2 Effect	-0.034	-0.026
					T3 Effect	-0.022	-0.022
	N	1751	1751	1751	N	1751	1751
Female Indirect	Treatment Effect	0.002	0.044*	-0.036*	Interaction Effect	-0.016	-0.014
	SE	0.021	0.023	0.019	SE	0.044	0.048
	Control Mean	0.164	0.142	0.175	Control Mean	0.144	0.144
					T1 or T2 Effect	0.01	0.051*
					T3 Effect	-0.026*	-0.026*
	N	1660	1660	1660	N	1660	1660

Data charts
Columns depict average levels of variable for each treatment arm and for different subgroups of the population.

Standard Errors
'Whiskers' at the top of the columns indicate standard errors (SE) of the mean. A large SE indicates a lot of uncertainty in the mean value, if they are tight, that represents a lot of certainty in the mean value.

Interactions

The table on the right shows the interaction effects between treatments.

T1*T3: interaction between T1 and T3

T2*T3: interaction between T2 and T3.

T1 or T2 effect: the effect of treatment T1 or T2 in the case in which T3=0

T3 effect: The T3 Effect term shows the effect of T3 when T1 or T2 is 0.

Level of significance

If the effect is significant at the 90% level we add one star (*), 95% level (**), and 99% level (***).

Results 1: Primary outcomes

IPV

This sub section includes results for all of the primary outcome measures relating to IPV, which are clustered together as follows:

- Women's experience of physical and/or sexual IPV (including specific and severe forms)
- Women's perceived changes in levels of IPV
- Women's experience of emotional abuse
- Men and boys' perpetration of physical and/or sexual IPV (including specific and severe forms)
- Men and boys' perpetration of emotional abuse.

The majority of primary outcome measures assessed changed in experienced and perpetrated violence between treatment and control by relying on ex ante balance between groups and accounting for baseline differences as controls. We also included two measures on women and men's perceptions of changes in levels of IPV.

Women's experiences of physical and/or sexual IPV

Survey data was gathered from 3,411 currently married or cohabiting women on the prevalence of physical and/or sexual IPV committed against them in the past three months.

Between 15% and 16% of women in the survey report having experienced some form of physical or sexual violence by an intimate partner in the previous three months, with rates very similar for direct beneficiaries across the different treatment arms. (Result 1)

Around 6% of women reported having experienced a severe form of IPV. When we look at a specific form of IPV, slapping or having objects thrown at them, we find between 7% and 9% of women report experiencing this form of IPV in the last three months, with marginally higher rates in the FI group. However, results from the embedded list experiment indicate significant underreporting of IPV. (Result 2). Analysis of responses from the list experiment suggests around 18% of FDs and 23% of FIs experienced being slapped or having something thrown at them at least once in the previous three months. This points to significant underreporting of the main outcome measure amongst women, meaning that actual prevalence rates were likely higher than reported.

There is no evidence that any of the interventions led to a reduction in women's experience of physical or sexual IPV amongst FDs (Result 1), and weak evidence of adverse effects on FIs. For FDs none of the treatment effects are statistically significant. Amongst FIs, we see some weak evidence of effects significant at the 90% level. FIs in T2 were significantly more likely to experience IPV, whereas FIs in T3 were significantly less likely to experience IPV. On the basis of the list experiment data, none of the estimated effects associated with any of the treatments are positive. In fact, a number are large and point in the 'wrong' direction - that is, consistent with an increase in experienced violence. (Result 2)

The absence of positive treatment effects is also supported by women's own perceptions of changes in levels of IPV with women on average seeing no change in levels of IPV over the past two years, with no differences across treatment and control. (Result 3)

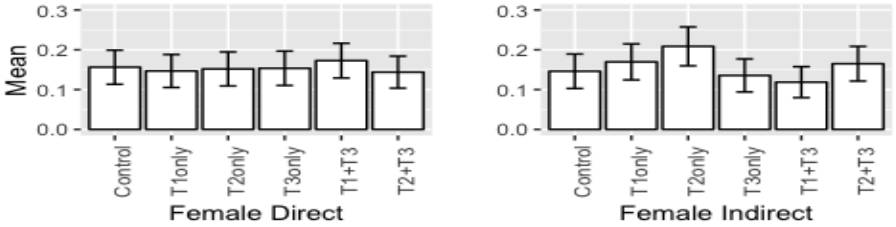
Women's experience of emotional IPV

In addition to physical and sexual IPV, women were asked about emotional abuse experienced from an intimate partner in the past three months. About 20% of women report experiencing emotional violence, with similar numbers among the direct and indirect beneficiaries. (Result 4)

As above, there is no evidence of treatment effects for FDs and weak evidence of treatment effects for FIs for this outcome. FIs in the T2 group report a relative increase in emotional violence - which could reflect greater sensitivity to emotional violence. On the other hand, we find some suggestive evidence that the T3 intervention led to a 25% decrease in emotional IPV for this group (statistically significant at the 95% level). (Result 4)

Result 1: Women's experience of any form of physical or sexual IPV

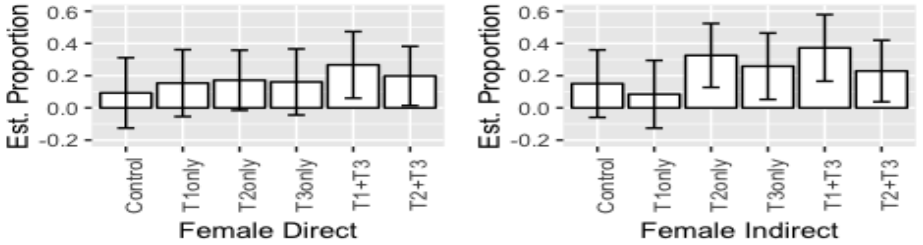
Description of measure: This measures the prevalence of physical or sexual IPV among married or cohabiting women (EIPV_ANY). It is coded as 1 if the respondent has experienced at least 1 of 7 forms of physical or sexual IPV in the last 3 months using EDV7-EDV13 in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	-0.006	-0.017	0.002	Interaction Effect	0.054	0.018
	SE	0.022	0.021	0.017	SE	0.042	0.042
	Control Mean	0.151	0.157	0.152	Control Mean	0.156	0.156
					T1 or T2 Effect	-0.034	-0.026
					T3 Effect	-0.022	-0.022
	N	1751	1751	1751	N	1751	1751
Female Indirect	Treatment Effect	0.002	0.044*	-0.036*	Interaction Effect	-0.016	-0.014
	SE	0.021	0.023	0.019	SE	0.044	0.048
	Control Mean	0.164	0.142	0.175	Control Mean	0.144	0.144
					T1 or T2 Effect	0.01	0.051*
					T3 Effect	-0.026*	-0.026*
	N	1660	1660	1660	N	1660	1660

RESULT 2: Women's experience of a specific form of physical IPV by women (LE)

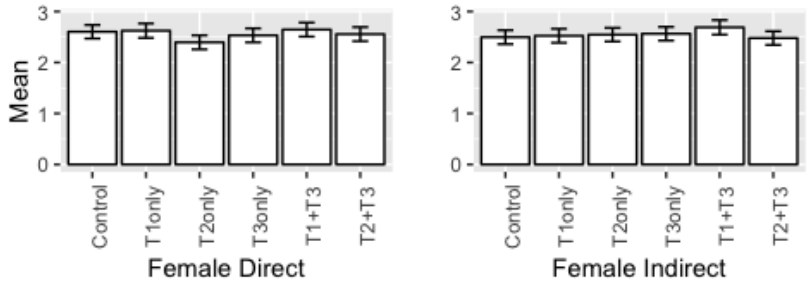
Description of measure: This measures the prevalence of a specific form of IPV among married or cohabiting women, as reported indirectly in a list experiment (EIPV_LE). See Box 2 in Part A for a description of the estimation of treatment effects from the list experiment. Measure based on lists of 3 or 4 statements using LE6 and LE7 in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment*Long	0.102	0.11	0.074	Interaction*Long	0.166	0.013
	SE	0.106	0.093	0.081	SE	0.214	0.182
	LongList Effect	0.181***	0.181***	0.181***	LongList Effect	0.181***	0.181***
					Treatment*Long	0.018	0.103
					T3*Longlist	0.014	0.014
	N	1754	1754	1754	N	1754	1754
Female Indirect	Treatment*Long	0.041	0.099	0.082	Interaction*Long	0.316	-0.088
	SE	0.1	0.09	0.081	SE	0.195	0.183
	LongList Effect	0.232***	0.232***	0.232***	LongList Effect	0.232***	0.232***
					Treatment*Long	-0.119	0.144
					T3*Longlist	0.007	0.007
	N	1660	1660	1660	N	1660	1660

RESULT 3: Women's perceptions of changes in IPV

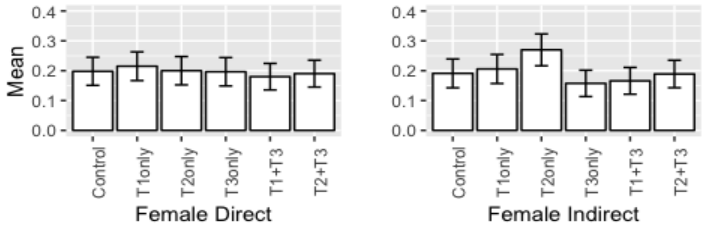
Description of measure: This measures the perceived change in experience of IPV among women who have been married/cohabiting for at least 2 years (EIPV_PERC). It is coded as on a scale of 1 to 5, where 1 indicates experiencing much less and 5 indicates experiencing much more IPV at present than 2 years ago using EDV15 in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	0.083	-0.064	0.031	Interaction Effect	0.109	0.301**
	SE	0.069	0.067	0.057	SE	0.14	0.138
	Control Mean	2.522	2.602	2.541	Control Mean	2.603	2.603
					T1 or T2 Effect	0.028	-0.217
					T3 Effect	-0.108	-0.108
	N	1752	1752	1752	N	1752	1752
Female Indirect	Treatment Effect	0.1	-0.008	0.007	Interaction Effect	0.081	-0.127
	SE	0.066	0.065	0.056	SE	0.141	0.134
	Control Mean	2.52	2.569	2.522	Control Mean	2.496	2.496
					T1 or T2 Effect	0.059	0.057
					T3 Effect	0.023	0.023
	N	1654	1654	1654	N	1654	1654

RESULT 4: Experience of any form of emotional IPV by women

Description of measure: This measures the prevalence of emotional abuse by an intimate partner among married or cohabiting women (EIPV_EMO). It is coded as 1 if the respondent has experienced at least 1 of 6 forms of emotional abuse from an intimate partner in the last 3 months using EDV1-EDV6 in the endline survey.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	0.004	-0.009	-0.016	Interaction Effect	-0.014	0.013
	SE	0.022	0.021	0.017	SE	0.045	0.042
	Control Mean	0.196	0.197	0.204	Control Mean	0.198	0.198
					T1 or T2 Effect	0.011	-0.015
					T3 Effect	-0.015	-0.015
	N	1751	1751	1751	N	1751	1751
Female Indirect	Treatment Effect	0.006	0.043*	-0.049**	Interaction Effect	0.043	-0.023
	SE	0.023	0.024	0.019	SE	0.045	0.046
	Control Mean	0.203	0.181	0.224	Control Mean	0.194	0.194
					T1 or T2 Effect	-0.016	0.055*
					T3 Effect	-0.056**	-0.056**
	N	1659	1659	1659	N	1659	1659

Men and boys' perpetration of physical and/or sexual IPV

Survey data was gathered from 1,157 currently married or cohabiting men on the prevalence of physical and/or sexual IPV perpetrated by them in the previous three months.[51]

A very small proportion of the MDs were married or cohabiting at the time of the survey reflecting their young age (n=189), whereas a significantly higher proportion of MIs were married or cohabiting (n=968). Approximately 10-14% of these married or cohabiting MD respondents reported engaging in physical or sexual IPV in the previous three months, slightly higher than the MIs (between 8-9%).

We do not detect a statistically significant reduction in levels of physical and/or sexual IPV perpetrated by MDs or MIs.[52] (Result 5). When we look at a specific form of IPV, slapping or throwing objects at an intimate partner, we find between 2% and 3% of men report this behaviour in the previous three months (less than half that reportedly experienced by women). Again there is no consistent evidence that this specific behaviour is affected by the treatments. (Result 6)

However, similar to findings on women's experiences, reports of slapping or throwing objects at women generated by the list experiment are considerably higher than those generated by the direct question, suggesting considerable underreporting by men. (Result 8). For the MIs for example the mean estimate is 21% and for the MDs the estimate is 34%. Using this measure, there is some weak indication of a decline in violence among MIs although it is not measured with great precision. (Result 8)

Fewer than 5% of married men report perpetration of severe forms of IPV (result 7). The SHG strengthening intervention appears to reduce the likelihood of married MIs having perpetrated a severe form IPV in the previous three months (significant at the 95% level). This result is important as there are sometimes concerns that efforts to strengthen women's groups can lead to backlash by men. Whilst the MI population did not include partners of FDs, from their own reports at least it appears that the SHG interventions (T1 and T2) have not led to an increase in levels of IPV within the wider community. (Result 7)

Men and boys' perpetration of emotional IPV

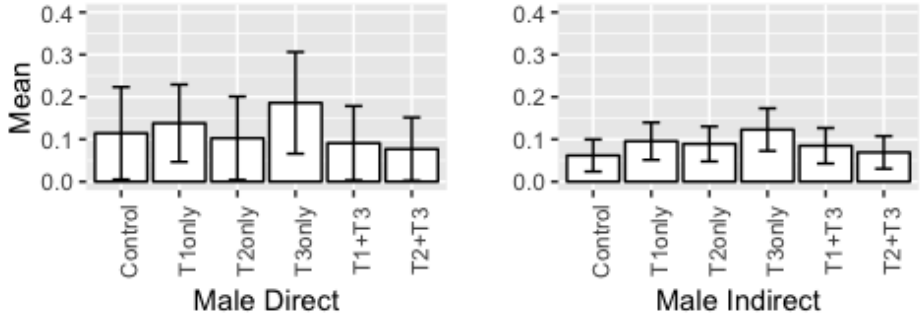
Rates of men and boys' perpetration of emotional violence is strikingly similar to the rates reporting having experienced of emotional violence, at around 20%. There are no significant effects discernible across the various treatment groups and control for this measure. (Result 9)

[51] Defined as women answering 'yes' to having experienced at least one of physical or sexual violence by an intimate partner at least once in the last 3 months.

[52] We estimate a drop in perpetration (not statistically significant) though this is measured with considerable noise; moreover it is largest in the T2 group, where the male direct beneficiaries did not actually receive any treatment.

RESULT 5: Perpetration of any form of physical or sexual IPV by men

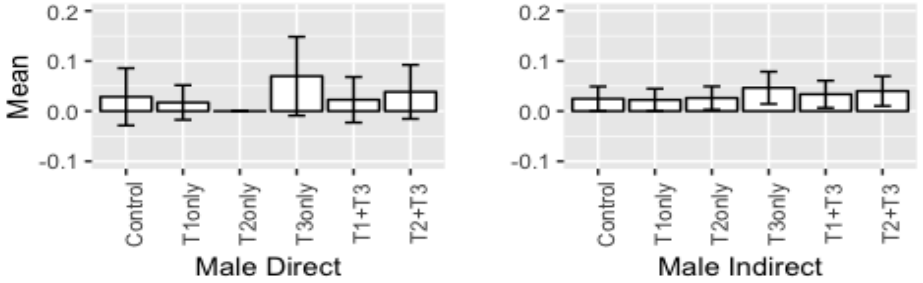
Description of measure: This measures the perpetration of physical or sexual IPV against women by married or cohabiting men (PIPV_ANY). It is coded as 1 if the respondent has perpetrated at least 1 of 7 forms of physical or sexual IPV in the last 3 months using PDV6-PDV12 in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Male Direct	Treatment Effect	0.025	-0.048	-0.032	Interaction Effect	0.04	0.335*
	SE	0.075	0.077	0.043	SE	0.115	0.172
	Control Mean	0.101	0.121	0.144	Control Mean	0.15	0.15
					T1 or T2 Effect	0.005	-0.212
					T3 Effect	-0.164	-0.164
	N	189	189	189	N	189	189
Male Indirect	Treatment Effect	-0.003	-0.019	0	Interaction Effect	-0.054	-0.074*
	SE	0.017	0.018	0.014	SE	0.035	0.038
	Control Mean	0.083	0.093	0.083	Control Mean	0.062	0.062
					T1 or T2 Effect	0.024	0.019
					T3 Effect	0.044	0.044
	N	968	968	968	N	968	968

RESULT 6: Perpetration of a specific form of IPV by men (directly reported)

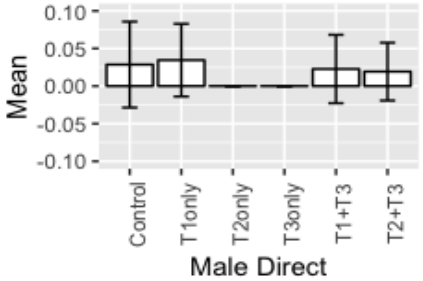
Description of measure: This measures the perpetration of a specific form of IPV against women by married or cohabiting men, as reported directly in a survey question (PIPV_SLAP). It is coded as 1 if the respondent has slapped or thrown something at an intimate partner in the last 3 months using PDV6 in the endline survey instrument.

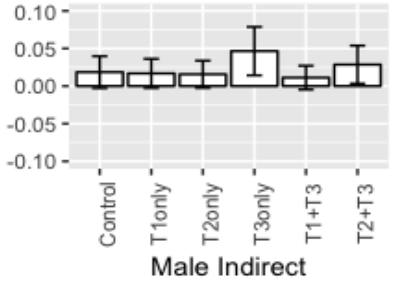


Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Male Direct	Treatment Effect	-0.019	-0.058	0.021	Interaction Effect	0.073	0.244**
	SE	0.062	0.067	0.037	SE	0.085	0.113
	Control Mean	0.037	0.032	0.021	Control Mean	0.05	0.05
					T1 or T2 Effect	-0.054	-0.177
					T3 Effect	-0.094	-0.094
	N	189	189	189	N	189	189
Male Indirect	Treatment Effect	-0.009	-0.005	0.012	Interaction Effect	0.001	0.001
	SE	0.013	0.013	0.01	SE	0.026	0.028
	Control Mean	0.034	0.034	0.025	Control Mean	0.028	0.028
					T1 or T2 Effect	-0.009	-0.005
					T3 Effect	0.012	0.012
	N	970	970	970	N	970	970

RESULT 7: Perpetration of a severe form of physical IPV by men

Description of measure: This measures the perpetration of severe forms of IPV against women by married or cohabiting men (PIPV_SEV). It is coded as 1 if the respondent has perpetrated at least 1 of 3 forms of severe physical IPV in the last 3 months using PDV8-PDV10 in the endline survey instrument.

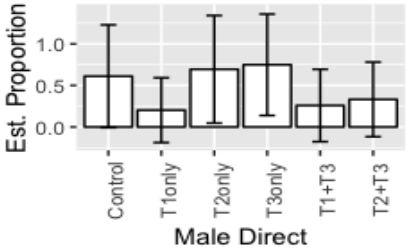


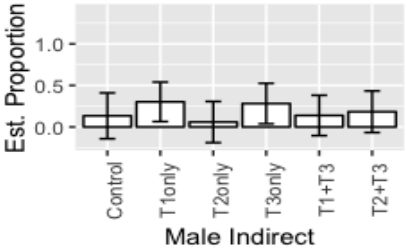


Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Male Direct	Treatment Effect	-0.042	-0.034	-0.037	Interaction Effect	-0.003	0.126
	SE	0.029	0.025	0.026	SE	0.055	0.077
	Control Mean	0.018	0.032	0.031	Control Mean	0.05	0.05
					T1 or T2 Effect	-0.041	-0.096
					T3 Effect	-0.079	-0.079
	N	189	189	189	N	189	189
Male Indirect	Treatment Effect	-0.025**	-0.013	0.01	Interaction Effect	-0.025	-0.009
	SE	0.011	0.012	0.009	SE	0.021	0.024
	Control Mean	0.03	0.025	0.019	Control Mean	0.021	0.021
					T1 or T2 Effect	-0.013**	-0.008
					T3 Effect	0.022	0.022
	N	970	970	970	N	970	970

RESULT 8: Perpetration of a specific form of IPV by men (list experiment)

Description of measure: This measures the perpetration of a specific form of IPV against women by married or cohabiting men, as reported indirectly in a list experiment (PIPV_LE). It is coded as the number of statements that a respondent indicates are true from a list of either 3 or 4 statements using LE8 and LE9 in the endline survey instrument.

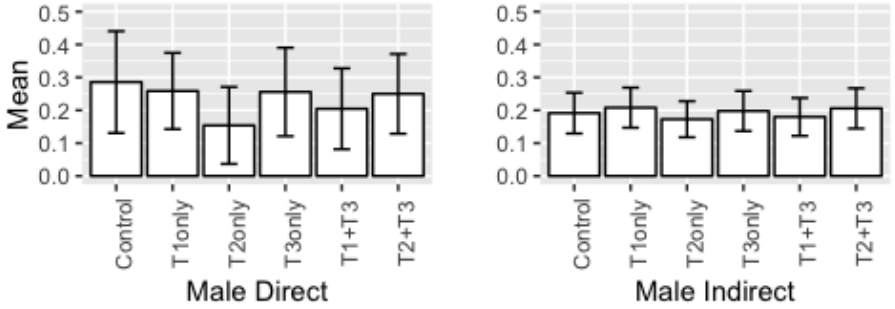




Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Male Direct	Treatment*Long	-0.314	-0.092	0.017	Interaction*Long	0.874	0.872
	SE	0.554	0.658	0.417	SE	1.248	1.409
	LongList Effect	0.344	0.344	0.344	LongList Effect	0.344	0.344
					Treatment*Long	-0.739	-0.517
					T3*Longlist	-0.653	-0.653
	N	189	189	189	N	189	189
Male Indirect	Treatment*Long	-0.017	-0.086	-0.055	Interaction*Long	0.051	-0.038
	SE	0.123	0.131	0.103	SE	0.244	0.252
	LongList Effect	0.212***	0.212***	0.212***	LongList Effect	0.212***	0.212***
					Treatment*Long	-0.043	-0.067
					T3*Longlist	-0.06	-0.06
	N	970	970	970	N	970	970

Result 9: Perpetration of emotional abuse against women by men

Description of measure: This measures the perpetration of emotional abuse against women by married or cohabitating men (EIPV_EMO). It is coded as 1 if the respondent has perpetrated at least 1 of 6 forms of emotional abuse in the last 3 months using PDV1A-PDV5 in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Male Direct	Treatment Effect	0.003	-0.048	0.065	Interaction Effect	0.008	0.367
	SE	0.093	0.084	0.059	SE	0.172	0.23
	Control Mean	0.239	0.242	0.247	Control Mean	0.3	0.3
					T1 or T2 Effect	-0.001	-0.226
					T3 Effect	-0.065	-0.065
	N	189	189	189	N	189	189
Male Indirect	Treatment Effect	-0.016	0.004	0.003	Interaction Effect	-0.022	0.043
	SE	0.026	0.026	0.02	SE	0.051	0.051
	Control Mean	0.189	0.191	0.187	Control Mean	0.186	0.186
					T1 or T2 Effect	-0.005	-0.017
					T3 Effect	-0.004	-0.004
	N	970	970	970	N	970	970

Findings from the qualitative data on IPV

Findings from the qualitative data are broadly supportive of the quantitative findings with limited evidence suggesting that levels of IPV have reduced as a result of any of the interventions.

In general, across all treatment groups, in FGDs women perceive IPV to be very common**.** This supports the conclusion, also to be drawn from the list experiment, that there was significant underreporting of personal experience of violence in the quantitative survey.

> *'If we do something wrong, then they beat us. If they commit a mistake, then they take it out on the family. Even on small mistakes....Small household mistakes. Women make mistakes, but don't men make them too? Whoever makes the mistake, the women gets beaten.'* (Woman DB, T2+T3 slum, Gwalior)
>
> *'Well if you talk about the entire area, then it is 10 out of 10! It happens all the time. Men come back from work in the evening, fully drunk and beat their wives'* (Woman DB, T2 slum, Indore)
>
> *'A little bit of beating/fighting happens at everyone's home. It happens at everyone's place. It doesn't happen at our place'.* (Woman DB, T1 slum, Bhopal)

When asked directly whether levels of IPV have changed, women and men direct beneficiaries– regardless of treatment type - report that IPV has either stayed the same or has reduced in their home slum, but increased in the city and country. Since this was the case in both intervention and control slums, this does not provide evidence that IPV levels have been affected by any of the treatment arms.

Furthermore, it likely indicates that the data may suffer from social desirability bias, with respondents wanting to protect the reputation of their home slum.[53]

> *'Nowadays people are very sensible and they are less violent because they know it affects our children. We also tell them, if you do this then it affects very badly in the future of your children'* (Woman DB, T2+T3 slum, Jabalpur)

> *'5 years ago there used to be a lot of violence in every home. Too much. Somewhere or other a man would be beating his wife with a stick. When the women became aware, they shared with one another and then it was controlled.'* (Woman DB, Control slum, Jabalpur)

However, the reasons given for why IPV is perceived to have reduced in their locality varied by intervention type, which may tell us something about levels of awareness and knowledge, as well as perceived benefits of SHG membership. Among FDs in SHG strengthening slums (T1), we found a common perception that physical and emotional IPV has decreased for members of SHGs, due to improvements in relationships with husbands. In particular, improved savings are perceived to have reduced household tensions, domestic disputes, and improved women's role and respect in the household. These improvements in household relationships are perceived by women as one of the most significant benefits of SHG membership.

> *'Earlier he used to yell at every small thing. But ever since we joined the group he does not object……. now he says I am not educated but he encourages me to learn. Asks me to learn and even teach them a few things. '* (Woman DB, T1 slum, Indore)

However, a few respondents report that some women SHG members faced an initial increase in IPV, which later declined as their husbands saw the benefits of their membership. This supports other qualitative findings from a recent systematic review of SHGs[54], though is not supported by the quantitative data.

Among FDs in SHG+VAW slums (T2), in addition to the reasons above, women also report that they feel more able to challenge their husbands' violent and controlling behaviour as they are more aware of forms of VAWG and their rights. However, if this were true we would expect to see an increase in self-reported experiences of IPV and VAWG in the quantitative survey – which we do not see. A few women even said that their husbands have changed as a result of them sharing what they had learned through the trainings. But in a few cases women said that IPV had decreased as they had learned to be quiet and not raise issues with their husbands, possibly indicating a shortcoming in the training.

Again it is important to note that these perceived changes in levels of IPV are not supported by the quantitative findings, and more likely indicate shifts in awareness and knowledge.

> *'Earlier we would not take care of the fact that our husbands come in the evening tired from work. The husband asks something, we didn't do it and thus conflicts begin. Then begin the verbal abuses. Now after the trainings, we shared our lessons with our husbands. So now even if we do something wrong – such as putting more salt in the vegetables than necessary, then he doesn't fight and let it go. So both of us become silent and it doesn't affect the children.'* (Woman DB, T2 slum, Indore)

Whilst there was considerable variation across FGDs, in SHG+VAWG (T2) women were generally more willing to talk about instances of IPV in their home slum than women in SHG only slums.[55] Although this increase is not associated with an increase

[53] A key finding at baseline was that reporting of IPV was significantly affected by respondents' desire to protect the reputation of their home slum. Similar to baseline findings, respondents across treatments and control were more willing to acknowledge and talk about IPV outside their home slum.

[54] Brody et al's (2016)

[55] Note from the baseline data, we found significant variation across slums of people's readiness to talk about IPV in their home slum, with some participants very reluctant to admit it happens in their locality related to a wider concern with the reputation of their colony in an environment where people living in slums can be viewed unfavourably.

in reporting, it may still signal an important first step in awareness and social norm change.

Violence and harassment in public spaces

This subsection includes all of the primary outcome measures relating to women's experience and men's perpetration of violence and harassment in public spaces.

Women's experience of VAWG in public spaces

Approximately 11% of women report experiencing violence and harassment in a public space in the previous three months. We find no evidence that that any of the interventions led to positive changes in this outcome with almost no differences across groups receiving treatment and control. (Result 10)

When we look at specific forms of public violence – touching, groping, stalking, flashing in public – levels of reported experience are much lower, around 5% for FDs, and 7% for FI report to have experienced at least one of these types of violence in the previous three months. Again there are almost no differences across groups receiving treatments and control. (Result 11)

Prevalence of these specific forms of violence and harassment calculated from the list experiment is considerably higher - around 12% - suggestive of underreporting. There is some variation across treatment groups using the list experiment measure but these are difficult to interpret. The biggest effect is for the life skills intervention (T3) for FDs. For this group there is large decline in reported experiences of violence and harassment in public; however there is no similar decline for the FIs, which is what one would expect if the effects are working through changes in the behaviour of men and boys. If the effect on FDs were consistent, one would also expect to see interactions when T3 is delivered in combination with T1 and T2 – however there are no such effects. Furthermore, there is no evidence of treatment effects when T3 works in combination with T1 and T2. (Result 12).

These findings on a lack of treatment effects is supported by women's perceptions of change in levels of VAWG in public spaces over the previous two years, with women reporting very little change with little variation across treatment groups. (Result 13).

Men and boys' perpetration of violence and harassment in public spaces

MDs are more likely to report perpetrating VAWG in public spaces than MIs (21-22% compared to 14-15%). As these samples have different age ranges this indicates that younger men are more likely to be perpetrators of public VAWG than older men.

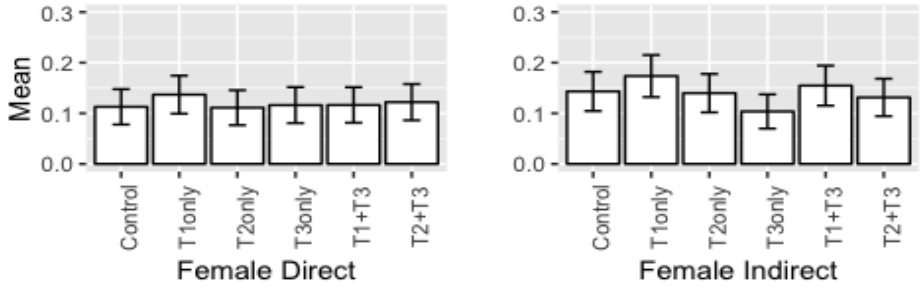
We do not find evidence of any treatment effects on this outcome**.** (Result 14)

Asked specifically about groping, stalking, flashing, about 6% of boys and men report engaging in these behaviours, with the rate for those in T3 being almost identical to those not in T3. Surprisingly, boys and men report higher levels of engagement in these activities in slums where T1 was implemented- i.e. where SHGs were strengthened but women were not receive the VAW intervention. (Result 15)

When we ask the same question through the list experiment, as with the other sensitive measures, we find that rates of perpetration of these specific forms of VAWG are much higher – over 20%. Again however there is no indication that the men and boys intervention (T3) reduced this activity. Moreover the increase associated with T1 above is not seen here, suggesting that that finding could have been a statistical anomaly. (Result 16)

RESULT 10: Experience of any form of harassment or violence in public by women

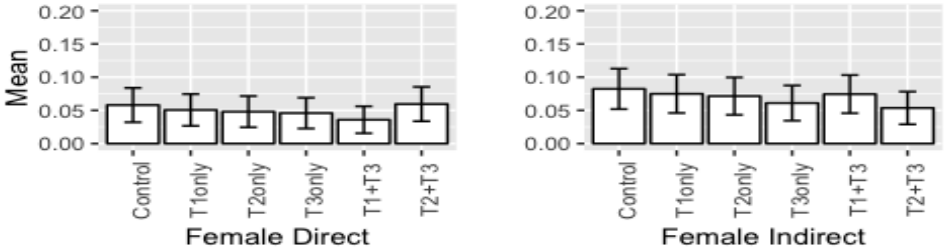
Description of measure: This measures the prevalence of public violence and harassment among women (EPV_ANY). It is coded as 1 if the respondent has experienced at least 1 of 5 forms of violence or harassment in a public space in the last 3 months using EPV1-5 in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	0.005	-0.006	0.003	Interaction Effect	-0.032	-0.002
	SE	0.017	0.017	0.013	SE	0.035	0.032
	Control Mean	0.116	0.121	0.12	Control Mean	0.113	0.113
					T1 or T2 Effect	0.021	-0.006
					T3 Effect	0.015	0.015
	N	1995	1995	1995	N	1995	1995
Female Indirect	Treatment Effect	0.03*	-0.002	-0.018	Interaction Effect	0.017	0.024
	SE	0.017	0.018	0.014	SE	0.035	0.035
	Control Mean	0.13	0.144	0.152	Control Mean	0.143	0.143
					T1 or T2 Effect	0.021*	-0.014
					T3 Effect	-0.031	-0.031
	N	1997	1997	1997	N	1997	1997

RESULT 11: Women's experience of a specific form of public harassment or violence (directly reported)

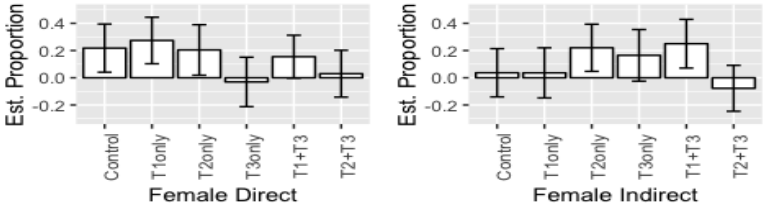
Description of measure: This measures the prevalence of a specific form of public harassment among women, as reported directly in a survey question (EPV_GROPE). It is coded as 1 if the respondent has been touched, groped, stalked or flashed by a man in a public place in the last 3 months using EPV2 in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	-0.015	-0.008	-0.005	Interaction Effect	-0.01	0.022
	SE	0.01	0.011	0.009	SE	0.021	0.021
	Control Mean	0.053	0.048	0.052	Control Mean	0.058	0.058
					T1 or T2 Effect	-0.01	-0.018
					T3 Effect	-0.009	-0.009
	N	1995	1995	1995	N	1995	1995
Female Indirect	Treatment Effect	0	-0.018	-0.013	Interaction Effect	0.014	-0.005
	SE	0.012	0.013	0.011	SE	0.023	0.024
	Control Mean	0.067	0.073	0.076	Control Mean	0.082	0.082
					T1 or T2 Effect	-0.007	-0.016
					T3 Effect	-0.016	-0.016
	N	1998	1998	1998	N	1998	1998

RESULT 12: Women's experience of harassment or violence in public (list experiment)

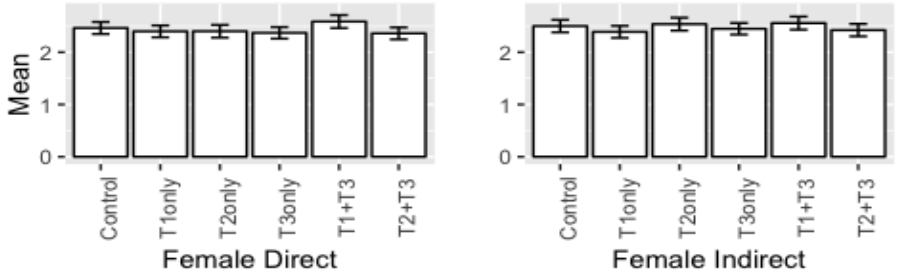
Description of measure: This measures the prevalence of a specific form of public harassment among women, as reported indirectly in a list experiment (EPV_LE). It is coded as the number of statements that a respondent indicates are true from a list of either 3 or 4 statements using LE2 and LE3 in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interaction		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment*Long	0.079	-0.031	-0.141**	Interaction *Long	0.065	0.075
	SE	0.075	0.086	0.065	SE	0.151	0.174
	LongList Effect	0.125***	0.125***	0.125***	LongList Effect	0.125***	0.125***
					Treatment *Long	0.046	-0.068
					T3*Longlist	-0.188**	-0.188**
	N	1996	1996	1996	N	1996	1996
Female Indirect	Treatment*Long	0.1	-0.002	-0.031	Interaction *Long	0.061	-0.314**
	SE	0.075	0.076	0.064	SE	0.153	0.151
	LongList Effect	0.105***	0.105***	0.105***	LongList Effect	0.105***	0.105***
					Treatment *Long	0.069	0.156
					T3*Longlist	0.054	0.054
	N	1998	1998	1998	N	1998	1998

RESULT 13: Women's perceptions of change in experience of public harassment or violence

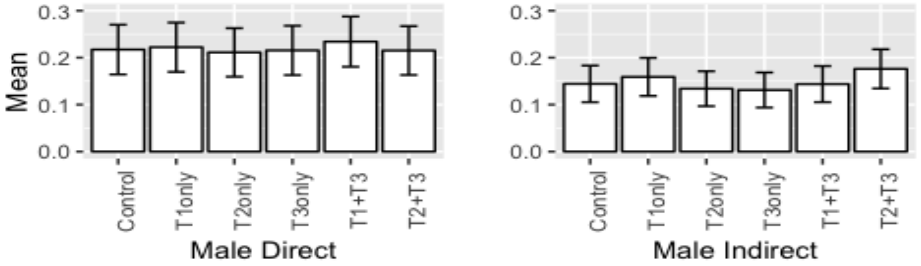
Description of measure: This measures the perceived change in experience of public violence or harassment among women (EPV_PERC). It is coded as on a scale of 1 to 5, where 1 indicates experiencing much less and 5 indicates experiencing much more public violence or harassment now than 2 years ago using EPV7 in the endline survey.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	0.082	-0.007	0.001	Interaction Effect	0.3**	0.041
	SE	0.057	0.054	0.044	SE	0.108	0.105
	Control Mean	2.398	2.455	2.421	Control Mean	2.463	2.463
					T1 or T2 Effect	-0.068	-0.028
					T3 Effect	-0.114	-0.114
	N	1995	1995	1995	N	1995	1995
Female Indirect	Treatment Effect	-0.014	0.012	-0.017	Interaction Effect	0.257**	-0.08
	SE	0.05	0.056	0.044	SE	0.097	0.116
	Control Mean	2.477	2.474	2.475	Control Mean	2.5	2.5
					T1 or T2 Effect	-0.143	0.052
					T3 Effect	-0.076	-0.076
	N	1997	1997	1997	N	1997	1997

RESULT 14: Perpetration of any form of public violence or harassment against women/girls by men

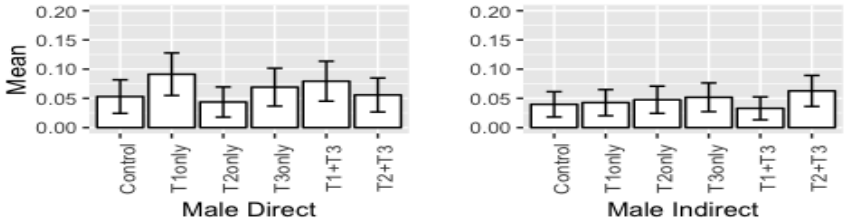
Description of measure: This measures the perpetration of violence or harassment against girls/women in public spaces by men (PPV_ANY). It is coded as 1 if the respondent has perpetrated at least 1 of 5 forms of violence or harassment against a girl or woman in the last 3 months in a public space using PPV1-PPV5 in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Male Direct	Treatment Effect	0.022	0.017	-0.006	Interaction Effect	0.002	-0.021
	SE	0.02	0.02	0.015	SE	0.039	0.037
	Control Mean	0.215	0.222	0.217	Control Mean	0.217	0.217
					T1 or T2 Effect	0.021	0.027
					T3 Effect	0.001	0.001
	N	1496	1496	1496	N	1496	1496
Male Indirect	Treatment Effect	0.018	0.025	-0.003	Interaction Effect	0.008	0.038
	SE	0.015	0.017	0.013	SE	0.031	0.034
	Control Mean	0.146	0.144	0.146	Control Mean	0.144	0.144
					T1 or T2 Effect	0.014	0.006
					T3 Effect	-0.018	-0.018
	N	1986	1986	1986	N	1986	1986

RESULT 15: Perpetration of a specific form of public violence or harassment against women/girls by men

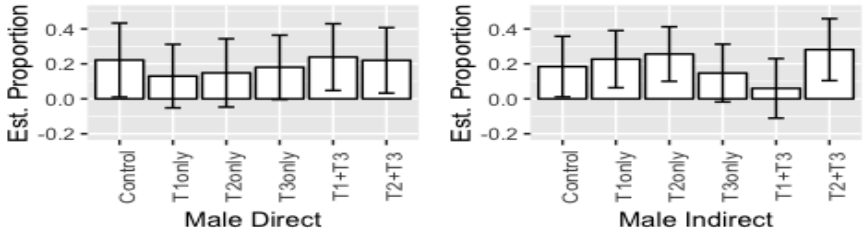
Description of measure: This measures the perpetration of a specific form of public harassment against girls/women by men, as reported directly in a survey question (PPV_GROPE). It is coded as 1 if the respondent has touched, groped, stalked or flashed a girl/woman in a public place in the last 3 months using PPV2 in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Male Direct	Treatment Effect	0.029**	-0.002	0.001	Interaction Effect	-0.034	-0.023
	SE	0.011	0.011	0.009	SE	0.023	0.023
	Control Mean	0.055	0.073	0.063	Control Mean	0.053	0.053
					T1 or T2 Effect	0.046**	0.009
					T3 Effect	0.021	0.021
	N	1498	1498	1498	N	1498	1498
Male Indirect	Treatment Effect	-0.005	0.014	0.002	Interaction Effect	-0.011	-0.002
	SE	0.01	0.01	0.009	SE	0.021	0.021
	Control Mean	0.051	0.042	0.043	Control Mean	0.04	0.04
					T1 or T2 Effect	0.001	0.015
					T3 Effect	0.006	0.006
	N	1988	1988	1988	N	1988	1988

RESULT 16: Perpetration of a specific form of public violence or harassment against women/girls by men (List Experiment)

Description of measure: This measures the perpetration of a specific form of public harassment against women by men, as reported indirectly in a list experiment (PPV_LE). It is coded as the number of statements that a respondent indicates are true from a list of either 3 or 4 statements using LE4 and LE5 in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Male Direct	Treatment*Long	-0.061	-0.069	0.048	Interaction*Long	0.289	0.116
	SE	0.096	0.089	0.078	SE	0.186	0.185
	LongList Effect	0.222***	0.222***	0.222***	LongList Effect	0.222***	0.222***
					Treatment*Long	-0.214	-0.13
					T3*Longlist	-0.089	-0.089
	N	1282	1282	1282	N	1282	1282
Male Indirect	Treatment*Long	-0.048	0.069	-0.008	Interaction*Long	-0.022	0.109
	SE	0.076	0.077	0.065	SE	0.152	0.152
	LongList Effect	0.226***	0.226***	0.226***	LongList Effect	0.226***	0.226***
					Treatment*Long	-0.037	0.012
					T3*Longlist	-0.038	-0.038
	N	1729	1729	1729	N	1729	1729

Findings from the qualitative data on violence and harassment in public spaces

The qualitative data broadly supports evidence from the list experiment, which indicates underreporting of violence and harassment in public. Violence and harassment in public is perceived as a common occurrence across all treatment arms, with some variations in respondents' willingness to talk about such incidents in their home slum. Similar to baseline findings, violence and harassment in public spaces is perceived to be more common against unmarried girls going to/from college than against older women.

In FGDs, some men and boys from T3 admitted that most men and boys are perpetrators and that, even if individually they do not approve, they may engage in 'eve-teasing' because of peer pressure, highlighting prevalent harmful norms around masculinity.

However, certain findings from the qualitative component contradict the quantitative evidence, with MDs from T3 perceiving levels of public VAWG to have decreased in their home slums. In particular, men and boys report that they are less likely to perpetrate violence as a result of their engagement with the programme – and in some cases that they also encourage their friends to be more respectful of women. However, this is not supported by the quantitative data and there is a high probability that this data suffers from agreement bias, since the beneficiaries were aware that reducing VAWG is a key objective of the intervention.

In some SHG+VAWG (T2) groups in Bhopal, public VAWG is also perceived to be decreasing due to action by the SHGs, which have reduced opportunities for men and boys' to harass women and girls with impunity. Some women reported that members of their SHGs now intervene in instances of public VAWG, or that men and boys no longer harass women and girls in front of SHG members. It is possible that such actions may not have succeeded in shifting overall levels of pubic VAWG in these slums if men and boys alter the places and times where they perpetrate.

> *'Whenever they see us, they keep shut if they are harassing someone but they don't leave that place and go'.* (Woman DB, T2 slum, Bhopal)

Similar to findings around IPV, we find some evidence that the T2 intervention has led to women feeling more able to talk about and discuss public VAWG. By comparison, women in T1 slums were mostly unwilling to talk about violence and harassment in public spaces taking place in their home slum, with a clear reluctance amongst some participants especially at the beginning of discussions to admit that public violence and harassment against women happens there at all. Similar to the baseline findings, this seems to be linked to the desire to protect the reputation of the slum and present it as a 'good neighbourhood'.

Results 2: Intermediate outcomes

The following section presents the endline results on the Programme's intermediate outcome measures, which are clustered around the following core components of the theory of change:

- Attitudes and social norms around VAWG
- Reporting VAW
- Women's economic empowerment
- Mobility and feelings of safety
- Taking actions to address VAWG.

Attitudes and social norms around VAWG

This first sub section includes results for all of the intermediate outcome measures relating to attitudes and social norms[56] around VAWG including:

- Gender equitable attitudes
- Attitudes towards IPV
- Attitudes towards public VAWG
- Descriptive norms around IPV
- Prescriptive norms around IPV
- Descriptive norms around public VAWG
- Prescriptive norms around public VAWG.

Gender equitable attitudes

In the endline survey, we include a modified version of the gender equitable men (GEM) scale [57] designed to capture attitudes on gender roles and women's empowerment. The modified scale included the following five statements:

- Women should be able to go anywhere they want whenever they want
- Doing the cooking, cleaning and washing are a woman's responsibility
- A man should have the final word about decisions in the home
- A man is justified in deciding who his wife can or cannot see or talk to
- It is a wife's obligation to have sex with her husband even if she does not feel like it.

Respondents were asked about the extent to which they agreed or disagreed with each statement, using a scale of 1-5 where 1 indicated 'strongly disagree' and 5 'strongly agree'.

Overall women score higher on this scale than men, averaging 3.7 for women and 3.3 to 3.4 for men on a scale of 1 to 5. These numbers suggests that on average women agree with statements supporting women's autonomy around issues such as how to spend time, who to see, and whether to have sex. However, they generally do not *strongly* agree with these statements. When we look at the breakdown by statement, interestingly two particular gender inequitable attitudes are prevalent amongst women and significantly worse than men's attitudes. Over half of all women (74%) agree or strongly agree that it is a wife's obligation to have sex with her

[56] Please note norms relating to reporting VAW are included in the reporting VAW subsection.

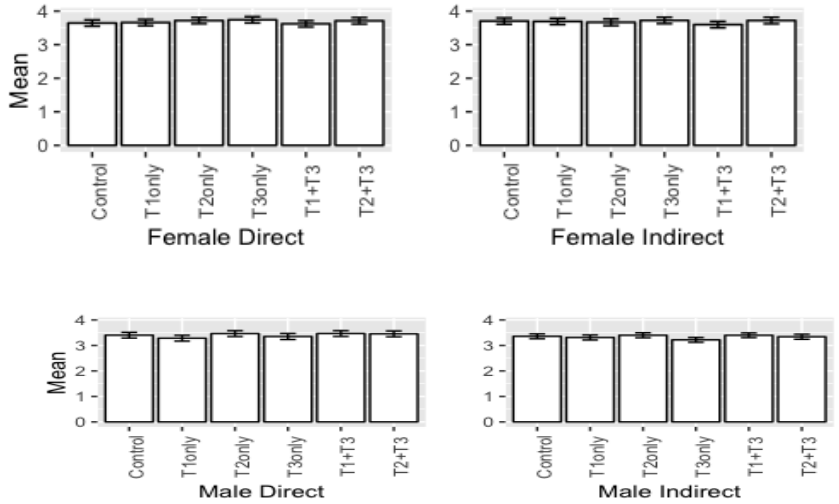
[57] The GEM scale was developed by Population Council/Horizons and Promundo to directly measure attitudes towards gender norms.

husband even if she does not feel like it and 68% of women believe that a man should have the final word about decisions in the home.[58]

There is no evidence of treatment effects on women or men's gender equitable attitudes. (Result 17)

RESULT 17: Attitudes (Modified GEM Scale)

Description of measure: This measures gender equitable attitudes towards mobility, gender roles, household decision-making, social relations and sexual relations (GEMSCALE). It is coded as the mean score of 5 attitudinal questions each measured on a scale from 1-5 where 1 indicates strong disagreement (agreement) and 5 indicates strong agreement (disagreement) with a gender equitable (inequitable) statement using AT1-5 in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	-0.034	0.024	0.022	Interaction Effect	-0.095	-0.079
	SE	0.056	0.05	0.043	SE	0.106	0.1

[58] The first of these findings highlights the challenges of measuring sexual IPV when forced sex within marriage is socially and culturally not acknowledged and supports likely underreporting of sexual IPV. Sexual violence within marriage is not currently recognised in Indian law. The Indian Penal Code prohibits sexual intercourse without consent, but there is an exception with respect to non-consent in the case of a husband and wife. This has been upheld in various judgements although under the India Domestic Violence Act, all forms of sexual abuse against wives is a civil wrong.

	Control Mean	3.703	3.666	3.673	Control Mean	3.644	3.644
					T1 or T2 Effect	0.014	0.064
					T3 Effect	0.08	0.08
	N	1996	1996	1996	N	1996	1996
Female Indirect	Treatment Effect	-0.059	-0.028	-0.021	Interaction Effect	-0.034	0.032
	SE	0.048	0.05	0.041	SE	0.097	0.103
	Control Mean	3.702	3.678	3.686	Control Mean	3.703	3.703
					T1 or T2 Effect	-0.043	-0.043
					T3 Effect	-0.02	-0.02
	N	1998	1998	1998	N	1998	1998
Male Direct	Treatment Effect	0.002	0.059	0.04	Interaction Effect	0.176*	-0.05
	SE	0.053	0.051	0.042	SE	0.106	0.102
	Control Mean	3.418	3.375	3.383	Control Mean	3.402	3.402
					T1 or T2 Effect	-0.087	0.085
					T3 Effect	-0.002	-0.002
	N	1498	1498	1498	N	1498	1498
Male Indirect	Treatment Effect	0.063	0.045	-0.025	Interaction Effect	0.17**	0.045
	SE	0.04	0.04	0.033	SE	0.084	0.079
	Control Mean	3.329	3.322	3.357	Control Mean	3.358	3.358
					T1 or T2 Effect	-0.023	0.023
					T3 Effect	-0.097	-0.097
	N	1988	1988	1988	N	1988	1988

Qualitative findings on gender equitable attitudes

The T2 and T3 curricula included a number of activities and messages designed to encourage beneficiaries to question inequitable gender norms, and in the case of men and boys, provide opportunities to trial new behaviours (for example through including cooking skills in gender trainings). From the qualitative data, there is some evidence of some of the messages in the curricula being reflected in attitudes of individuals that received T2 and T3.

In the T2 groups, there is some evidence that individual FDs are starting to question gender inequitable norms. However, despite these more positive attitudes, there is little evidence of wider shifts on gender roles or individual behaviours. As found at baseline, social norms in urban slums in Madhya Pradesh dictate that a woman's primary role is to be a good housewife and to undertake all domestic work *inside* the home, whilst a man's primary role is to be the breadwinner and to undertake all work *outside* of the household to support the family.

In the men and boys groups (T3), there appears to be greater agreement that men/boys should support women in household work than amongst those in control areas. And in a few cases men and boys report trialling these new behaviours in their home. The qualitative data highlights the influence of social expectations of peers on such behaviours. Respondents report that they previously felt 'ashamed' and 'embarrassed' to do such 'women's work', and that they previously feared being mocked by their friends, whereas now they have that social approval of other members of the group.

These findings for T3 MDs contrast with control MDs, where very strict norms around gender roles persist. Such deeply entrenched norms are reflected in the small number of men and boys in T3 who report such changes in their own behaviour.

Box 4: Qualitative exercises to explore gender equitable attitudes

The FGDs included an exercise to explore gender roles, norms and attitudes. Participants were asked to discuss and agree within the group whether they agreed or disagreed with statements around gender roles and stereotypes. Furthermore, gender (in)equitable attitudes which emerged at other points during discussions were coded and findings summarised in this section.

> *'Men cooking food in a hotel will not be ashamed of doing such work but if the same work we do at home, we will told that we are doing household work which is wrong. If a man cooks for thousand customers in hotel, it is remarked as right, and if the same man cooks at home for 4 members of his family, then he is tagged as not a man which is wrong.'* (Boy DB, Bhopal,T2+T3 slum)

> *'I didn't help at house earlier and whenever I was asked to do something I would say what will my friends say about me? But now I have started sweeping and cleaning my house and help my mother in her work'.* (Boy DB, Bhopal, T2+T3 slum)

Attitudes towards IPV and VAWG in public spaces

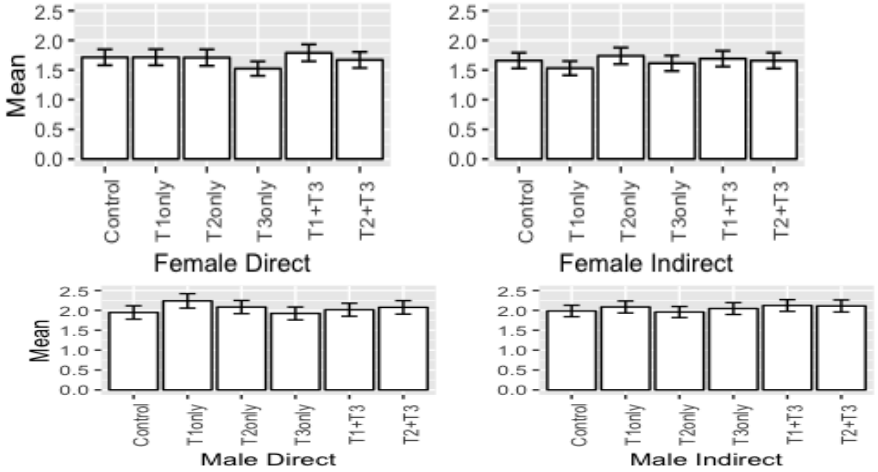
Respondents were asked whether they agree with the statement that a man should beat his wife if she disobeys him. Most respondents disagree with the statement though most not strongly. Men are more likely to agree - indeed about 26% of men agree or strongly agree with the statement compared to 16% of women.

There is little evidence of treatment effects on this outcome. There is no evidence that T1 or T2 increased women's disagreement with this statement. T3 boys and men are less likely to agree however. While this appears as evidence for a positive effect it is offset by the fact the boys and men report *stronger* support for beatings in T1 and T3 areas. (Result 18)

Most respondents disagree that harassing women in public is harmless fun, though men are more likely to agree. Just over 10% of men and boys agreed this behaviour is harmless fun compared to less than 5% of women. Attitudes, while different between gender groups, are no different between treatment arms indicating that the interventions had no effect in attitudes towards public VAWG. (Result 19)

RESULT 18: Individual attitudes on IPV perpetration

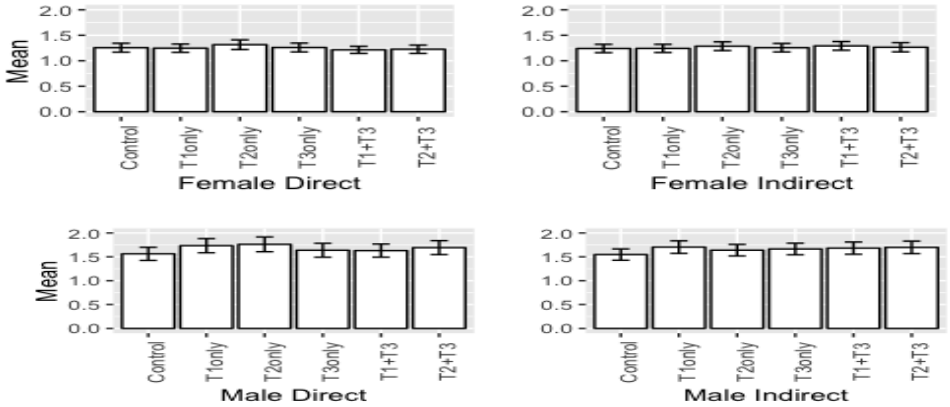
Description of measure: This measures the extent to which a respondent agrees that a man should beat his wife if she disobeys him (ATT_PIPV). It is coded as on a scale of 1-5 where 1 is strongly disagree and 5 is strongly agree using SN15 in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	0.087	0.019	-0.042	Interaction Effect	0.268**	0.233*
	SE	0.069	0.063	0.05	SE	0.133	0.123
	Control Mean	1.654	1.686	1.712	Control Mean	1.713	1.713
					T1 or T2 Effect	-0.048	-0.098
					T3 Effect	-0.21	-0.21
	N	1996	1996	1996	N	1996	1996
Female Indirect	Treatment Effect	-0.054	0.035	0.022	Interaction Effect	0.143	0.006
	SE	0.065	0.065	0.052	SE	0.138	0.13
	Control Mean	1.667	1.623	1.642	Control Mean	1.659	1.659
					T1 or T2 Effect	-0.126	0.032
					T3 Effect	-0.028	-0.028
	N	1998	1998	1998	N	1998	1998
Male Direct	Treatment Effect	0.167**	0.145**	-0.118**	Interaction Effect	-0.137	0.024
	SE	0.073	0.067	0.056	SE	0.144	0.132
	Control Mean	2.008	2.032	2.091	Control Mean	1.947	1.947
					T1 or T2 Effect	0.235**	0.133**
					T3 Effect	-0.08**	-0.08**
	N	1498	1498	1498	N	1498	1498
Male Indirect	Treatment Effect	0.074	0.022	0.046	Interaction Effect	0.025	0.126
	SE	0.057	0.06	0.048	SE	0.117	0.12
	Control Mean	2.025	2.06	2.009	Control Mean	1.985	1.985
					T1 or T2 Effect	0.062	-0.041
					T3 Effect	-0.005	-0.005
	N	1988	1988	1988	N	1988	1988

RESULT 19: Individual attitudes on public harassment perpetration

Description of measure: This measures the extent to which a respondent agrees that sexually harassing women is harmless fun (ATT_PPV). It is coded as on a scale of 1-5 where 1 is strongly disagree and 5 is strongly agree using SN16 in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	-0.042	-0.003	-0.038	Interaction Effect	-0.12*	-0.082
	SE	0.032	0.034	0.029	SE	0.065	0.067
	Control Mean	1.264	1.244	1.273	Control Mean	1.256	1.256
					T1 or T2 Effect	0.018	0.038
					T3 Effect	0.03	0.03
	N	1996	1996	1996	N	1996	1996
Female Indirect	Treatment Effect	-0.016	-0.009	0.021	Interaction Effect	-0.014	-0.017
	SE	0.034	0.038	0.028	SE	0.07	0.073
	Control Mean	1.262	1.258	1.257	Control Mean	1.241	1.241
					T1 or T2 Effect	-0.01	0
					T3 Effect	0.032	0.032
	N	1998	1998	1998	N	1998	1998
Male Direct	Treatment Effect	0.031	0.117*	-0.05	Interaction Effect	-0.219*	-0.174
	SE	0.068	0.067	0.052	SE	0.128	0.13
	Control Mean	1.665	1.642	1.688	Control Mean	1.563	1.563
					T1 or T2 Effect	0.141	0.204*
					T3 Effect	0.082	0.082
	N	1498	1498	1498	N	1498	1498
Male Indirect	Treatment Effect	0.05	0.034	0.026	Interaction Effect	-0.153	-0.06
	SE	0.054	0.055	0.047	SE	0.116	0.115
	Control Mean	1.638	1.65	1.631	Control Mean	1.547	1.547
					T1 or T2 Effect	0.127	0.064
					T3 Effect	0.097	0.097
	N	1988	1988	1988	N	1988	1988

Social norms around IPV

The survey included questions to measure both descriptive norms (i.e. shared beliefs about what is typical behaviour within a reference group – in this case other women and men in their social network) and prescriptive norms (shared beliefs about what is appropriate behaviour within their reference group).

Men and women believe that on average almost a third of men (approx. 30%) beat their wives if they disobey them. This is almost double the prevalence of physical or sexual IPV reported by women, and three times the reported prevalence of a specific

form of IPV (results tables 1 and 2), demonstrating how people's perceptions of levels of IPV are higher than reported levels. (Result 20).

There is limited and patchy evidence of treatment effects on descriptive norms. FDs in T3 treatment arms report slightly lower perceived prevalence of IPV. However, there is not a similar effect for FIs, which one would expect to see if there was a general trend at the community level. MIs in T2 report slightly higher perceived prevalence of IPV. Whilst this could indicate that activities of women's SHGs in T2 slums may have increased perceptions that IPV is a typical behaviour in the slum, as there is no similar effect for MDs, this seems unlikely.

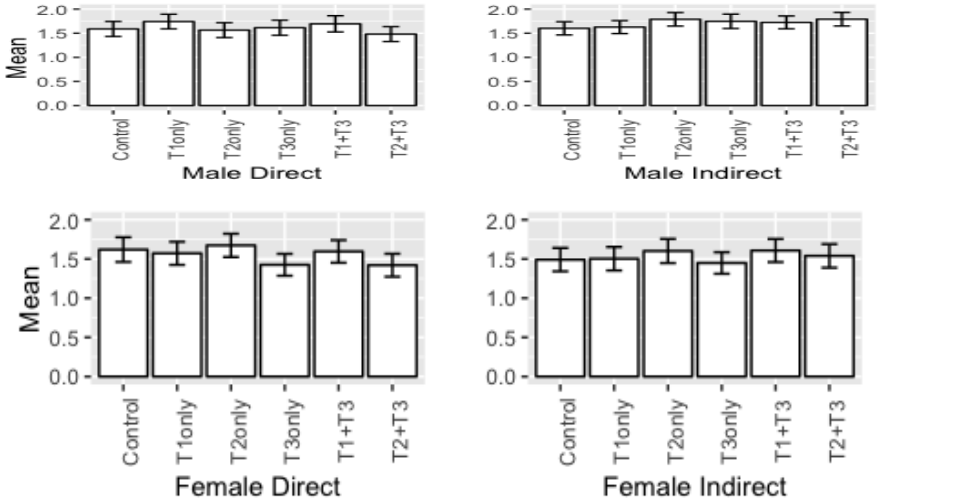
Findings on prescriptive norms about men's approval are similar to individual attitudes with approximately 15% of women and men believing that most other men are broadly approving of IPV. Women are slightly more likely than men to believe that men would be approving of IPV perpetration by other men. Fewer respondents believe that women are broadly approving of IPV (between 6 and 7%). (Result 21 and 22).

The findings highlight that descriptive norms are more pervasive than prescriptive norms. That is, beliefs about what proportion of men *would* use violence is higher than the proportion that report they *should* use violence in these circumstances.

There is some variability across treatment groups in prescriptive norms, though not in a way consistent with programme goals. Since indeed most men claim not to be supportive of these actions, the programme might have resulted in shared beliefs that more men are less approving of IPV. This is in general not the case however. There are no differences for FDs associated with T1 or T2 at the 95% level. The strongest effects, surprisingly, are for the women (potential) direct beneficiaries in T3 who are slightly less likely to believe that most women and men approve of IPV (statistically significant at the 95% level). However, since there is no such effect on FIs, nor evidence of any interactions when T3 is delivered in combination with T1 or T2, this is suggestive of null effects overall.

RESULT 20: Descriptive norms on IPV perpetration

Description of measure: This measures the perceived prevalence of IPV perpetration within a respondent's social network (DNORM_PIPV). It is coded as the number (out of 5 married men) that a respondent believes would beat their wives using SN3 in the endline survey instrument.

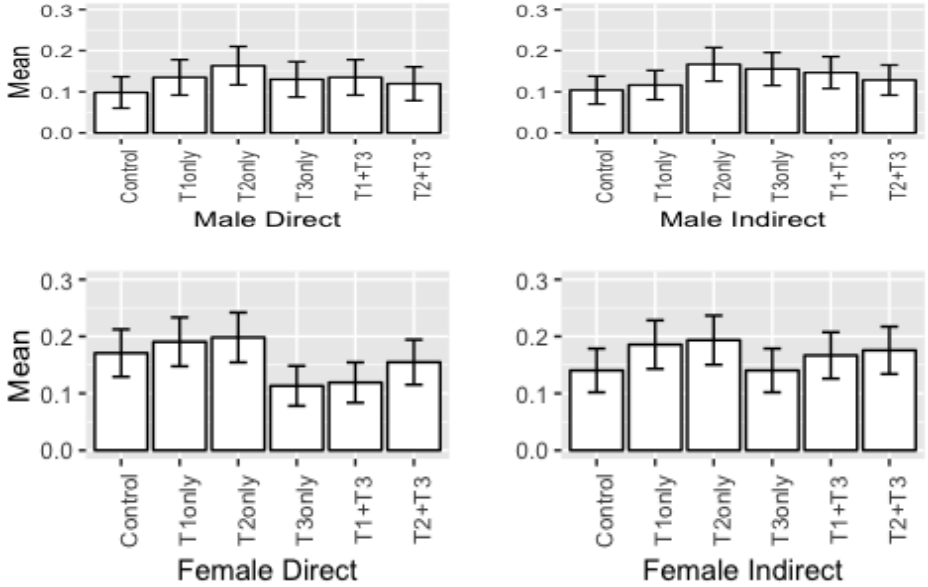


Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	0.043	-0.003	-0.143**	Interaction Effect	0.135	-0.098
	SE	0.071	0.071	0.056	SE	0.139	0.143
	Control Mean	1.534	1.553	1.621	Control Mean	1.619	1.619
					T1 or T2 Effect	-0.024	0.046
					T3 Effect	-0.155**	-0.155**
	N	1996	1996	1996	N	1996	1996
Female Indirect	Treatment Effect	0.05	0.055	0.007	Interaction Effect	0.108	0.009

	SE	0.066	0.068	0.055	SE	0.136	0.138
	Control Mean	1.52	1.513	1.532	Control Mean	1.491	1.491
					T1 or T2 Effect	-0.004	0.05
					T3 Effect	-0.032	-0.032
	N	1998	1998	1998	N	1998	1998
Male Direct	Treatment Effect	0.11	-0.071	-0.061	Interaction Effect	0.001	-0.08
	SE	0.071	0.066	0.056	SE	0.142	0.131
	Control Mean	1.562	1.661	1.632	Control Mean	1.588	1.588
					T1 or T2 Effect	0.11	-0.03
					T3 Effect	-0.035	-0.035
	N	1494	1494	1494	N	1494	1494
Male Indirect	Treatment Effect	0.014	0.136**	0.052	Interaction Effect	0.01	-0.146
	SE	0.062	0.064	0.052	SE	0.121	0.121
	Control Mean	1.733	1.675	1.673	Control Mean	1.601	1.601
					T1 or T2 Effect	0.01	0.209**
					T3 Effect	0.098	0.098
	N	1986	1986	1986	N	1986	1986

RESULT 21: Prescriptive norms on IPV perpetration (Men)

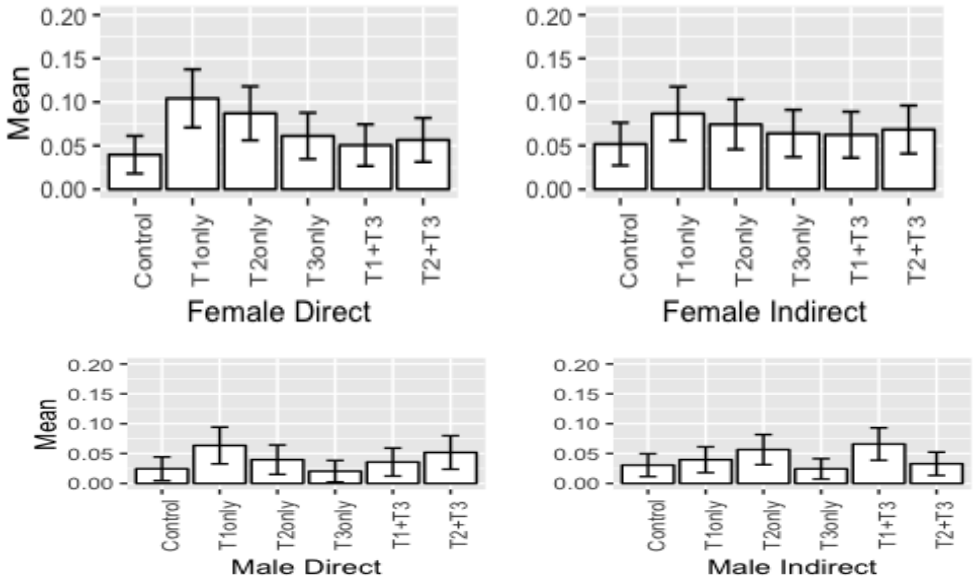
Description of measure: This measures the perceived approval of IPV perpetration within a respondent's social network of men (PNORM_PIPV_M). It is coded as a 1 if a respondent thinks that most out of 5 married men would approve of a man beating his wife if she disobeyed him using SN4A in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	0.003	0.021	-0.06***	Interaction Effect	-0.029	0.007
	SE	0.019	0.018	0.016	SE	0.037	0.037
	Control Mean	0.159	0.148	0.187	Control Mean	0.171	0.171
					T1 or T2 Effect	0.017	0.017
					T3 Effect	-0.053***	-0.053***
	N	1996	1996	1996	N	1996	1996
Female Indirect	Treatment Effect	0.036*	0.037**	-0.012	Interaction Effect	-0.018	-0.031
	SE	0.021	0.019	0.017	SE	0.042	0.038
	Control Mean	0.163	0.158	0.173	Control Mean	0.14	0.14
					T1 or T2 Effect	0.045*	0.052**
					T3 Effect	0.004	0.004
	N	1998	1998	1998	N	1998	1998
Male Direct	Treatment Effect	0.022	0.035*	-0.008	Interaction Effect	-0.031	-0.064
	SE	0.022	0.021	0.017	SE	0.045	0.041
	Control Mean	0.128	0.125	0.133	Control Mean	0.098	0.098
					T1 or T2 Effect	0.037	0.067*
					T3 Effect	0.024	0.024
	N	1496	1496	1496	N	1496	1496
Male Indirect	Treatment Effect	0.004	0.022	0.009	Interaction Effect	-0.035	-0.082**
	SE	0.015	0.015	0.012	SE	0.03	0.031
	Control Mean	0.139	0.131	0.129	Control Mean	0.104	0.104
					T1 or T2 Effect	0.021	0.063
					T3 Effect	0.048	0.048
	N	1987	1987	1987	N	1987	1987

RESULT 22: Prescriptive norms on IPV perpetration (Women)

Description of measure: This measures the perceived approval of IPV perpetration within a respondent's social network of women (PNORM_PIPV_F). It is coded as 1 if a respondent thinks that most out of 5 married women would approve of a man beating his wife if she disobeyed him using SN5A in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	0.019	0.012	-0.022**	Interaction Effect	-0.084***	-0.047**
	SE	0.011	0.011	0.01	SE	0.023	0.023
	Control Mean	0.061	0.064	0.077	Control Mean	0.04	0.04
					T1 or T2 Effect	0.061	0.035
					T3 Effect	0.023**	0.023**
	N	1996	1996	1996	N	1996	1996
Female Indirect	Treatment Effect	0.01	0.004	-0.004	Interaction Effect	-0.048**	-0.015
	SE	0.012	0.011	0.01	SE	0.024	0.024
	Control Mean	0.065	0.066	0.071	Control Mean	0.052	0.052
					T1 or T2 Effect	0.034	0.011
					T3 Effect	0.017	0.017
	N	1998	1998	1998	N	1998	1998
Male Direct	Treatment Effect	0.021*	0.023**	-0.004	Interaction Effect	-0.039	0.016
	SE	0.011	0.012	0.01	SE	0.024	0.023
	Control Mean	0.034	0.036	0.043	Control Mean	0.024	0.024
					T1 or T2 Effect	0.04*	0.015**
					T3 Effect	0.003	0.003
	N	1498	1498	1498	N	1498	1498
Male Indirect	Treatment Effect	0.025**	0.018**	-0.002	Interaction Effect	0.03	-0.015
	SE	0.01	0.009	0.008	SE	0.02	0.017
	Control Mean	0.036	0.04	0.042	Control Mean	0.031	0.031
					T1 or T2 Effect	0.01**	0.025**
					T3 Effect	-0.007	-0.007
	N	1988	1988	1988	N	1988	1988

Qualitative findings on social norms around IPV

The qualitative findings are broadly supportive of the survey findings with limited evidence of change in either descriptive or prescriptive norms around IPV, highlighting the pervasive nature of harmful norms driving VAWG in urban Madhya Pradesh.

FGDs included a number of exercises and discussion topics, which were designed to explore social norms around IPV, and the nuances and contexts in which norms may operate.[59] IPV is perceived by most groups (treatment and control) as common. Overall, the women's groups (both T1,T2 and control) perceive IPV to be more common than the men and boys (both treatment and control). MDs (both T3 and control) were more reluctant to talk about IPV, and there was a shared belief that this is a behaviour which they know little about as it happens in private between married couples.

> *'Well if you talk about the entire area, then it is 10 out of 10! It happens all the time. Men come back from work in the evening, fully drunk and beat their wives'* (Woman DB, T2 slum, Indore)

> *'A little bit of beating/fighting happens at everyone's home. It happens at everyone's place. It doesn't happen at our place'.* (Woman DB, T1, Bhopal)

Although across the FGDs (both control and treatments) women voiced general views broadly against IPV, there is also varying degrees of tolerance shown across the FGDs. There appears to be little difference in T1 and T2 groups in tolerance of IPV – compared with control and baseline findings. In most T1 and T2 groups, despite a few individual women believing that women should never have to tolerate abuse, there is a general social expectation that women should tolerate 'minor' or 'small' incidents of violence (such as minor beatings) 'up to a point', which happens in every home in order to keep the family together. Violence is justified when the wife is seen as being at fault – for example for burning food or not putting enough salt in food. This tendency to minimise violence is similar to baseline findings and signals the extent to which IPV is normalised, and still seen as a family matter, as well as the underlying assumption that IPV is often the fault of the survivor or both parties are to blame. The perception that most women will tolerate IPV is also strongly linked to perceptions that there are no alternatives for women – given inadequate local response and cultural norms around the importance of marriage.

> *'See the thing is – most of the fights happen when you pour too much salt in the vegetables. So if he doesn't like the food, he argues with you, or if he beats you up a little bit at that time, then you should also tolerate it, thinking alright, it was my fault; the food was not good.'* (Woman DB, T2 slum, Jabalpur)

> *'If in some cases he hits us 3-4 times it is no big deal. If it happens a lot, then people feel troubled.'* (Woman DB, T2 slum, Gwalior)

> *'We should tolerate. If we start arguing and fighting then the whole locality will hear…..She should tolerate it madam. If she will argue, he will beat her more and if she goes to court, it is her family who will be at loss….It should be tolerated if it is not much. She should not tolerate after a point.'* (Woman DB, T1 slum, Gwalior)

Underlying and sustaining IPV, and presenting significant challenges to VAWG response is a **harmful social norm that domestic violence is a family matter**. This norm persists in both treatment and control slums and serves to sustain violence and acceptance of IPV.

Social norms around public VAWG

The survey also included questions to measure descriptive and prescriptive norms around public VAWG.

Overall respondents report that they believe about 20% of married men in their social network would engage in harassment of women in public spaces. This is much higher than reported individual behaviours, as well as individual attitudes suggesting

[59] Including a video vignette depicting a case of IPV. See annex 4 for qualitative FGD guide.

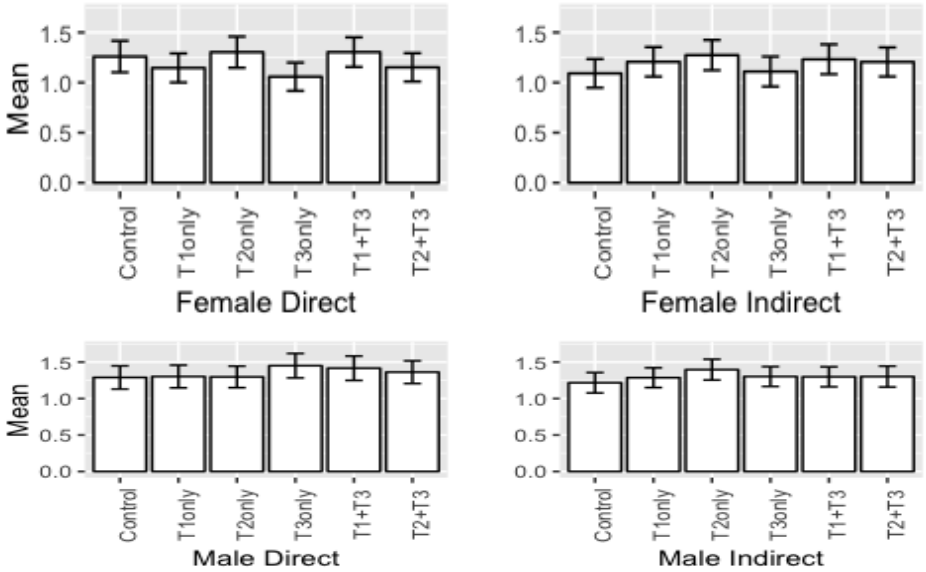
that respondents believe that men engage in this behaviour at higher rates than reported or individually find acceptable. (Result 23)

However, when we look at prescriptive norms, perceptions of social approval are much lower with most respondents believing that most married men and women in their social network *do not* approve of harassment in public places (between 90-95% of respondents). These findings may indicate that descriptive norms are a bigger driver of VAWG in public spaces than prescriptive norms – that is that men and boys may perpetrate VAWG in public spaces because they think other men in their social network do. (Result 24 and 25)

There is limited evidence of treatment effects on either descriptive or prescriptive norms, though these effects are mostly weak and not consistent with Programme goals. We see a small weak effect in T3 with MDs believing slightly more men in their social network would perpetrate IPV. FDs in T2 and FIs in T3 are slightly more likely to believe that most men in their social network approve of IPV, whereas FDs in T2 are slightly less likely to believe that most women in their social network approve of IPV. (Results 24 and 25)

RESULT 23: Descriptive norms on public harassment perpetration

Description of measure: This measures the perceived prevalence of public sexual harassment within a respondent's social network (DNORM_PPV). It is coded as the number (out of 5 married men) that a respondent believes would harass women on the street for fun using SN6 in the endline survey instrument.

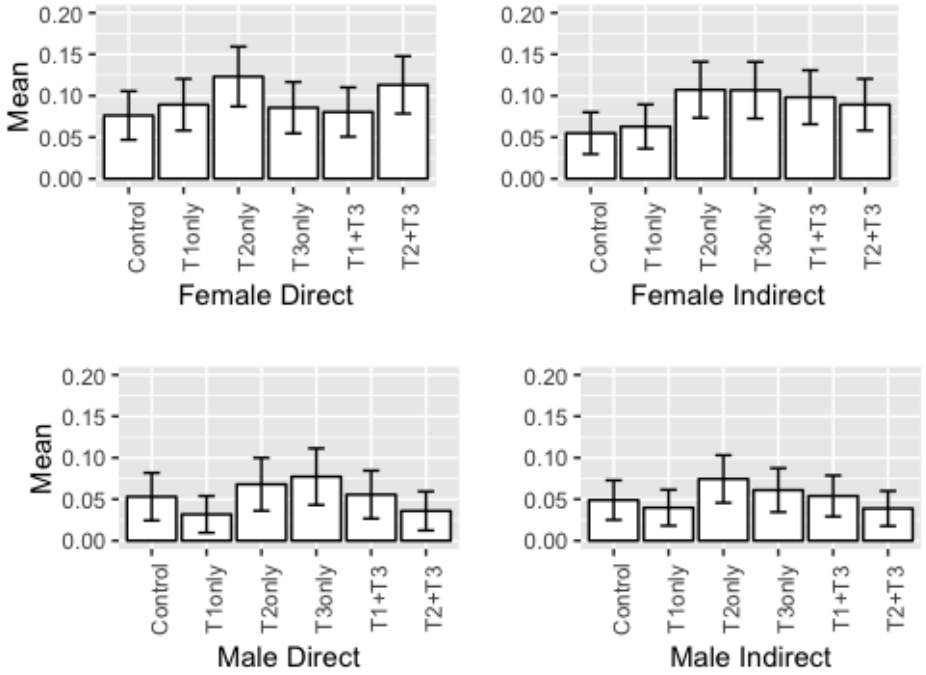


Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	0.021	0.032	-0.043	Interaction Effect	0.228*	0.068
	SE	0.06	0.059	0.049	SE	0.122	0.112
	Control Mean	1.193	1.192	1.236	Control Mean	1.259	1.259
					T1 or T2 Effect	-0.093	-0.003
					T3 Effect	-0.143	-0.143
	N	1996	1996	1996	N	1996	1996

Female Indirect	Treatment Effect	0.031	0.071	0.024	Interaction Effect	-0.046	-0.014
	SE	0.06	0.061	0.051	SE	0.123	0.124
	Control Mean	1.171	1.161	1.191	Control Mean	1.091	1.091
					T1 or T2 Effect	0.054	0.078
					T3 Effect	0.044	0.044
	N	1998	1998	1998	N	1998	1998
Male Direct	Treatment Effect	-0.039	-0.054	0.107*	Interaction Effect	-0.077	-0.035
	SE	0.075	0.071	0.06	SE	0.154	0.145
	Control Mean	1.35	1.365	1.297	Control Mean	1.29	1.29
					T1 or T2 Effect	-0.001	-0.037
					T3 Effect	0.144*	0.144*
	N	1497	1497	1497	N	1497	1497
Male Indirect	Treatment Effect	0.002	0.09	-0.023	Interaction Effect	-0.104	-0.14
	SE	0.07	0.076	0.057	SE	0.136	0.151
	Control Mean	1.305	1.277	1.301	Control Mean	1.217	1.217
					T1 or T2 Effect	0.054	0.16
					T3 Effect	0.059	0.059
	N	1985	1985	1985	N	1985	1985

RESULT 24: Prescriptive norms on public harassment perpetration (Men)

Description of measure: This measures the perceived approval of public sexual harassment perpetration within a respondent's social network of men (PNORM_PPV_M). It is coded as 1 if a respondent thinks that most out of 5 married men would approve of a man sexually harassing a woman on the street using SN7 in the endline survey instrument.

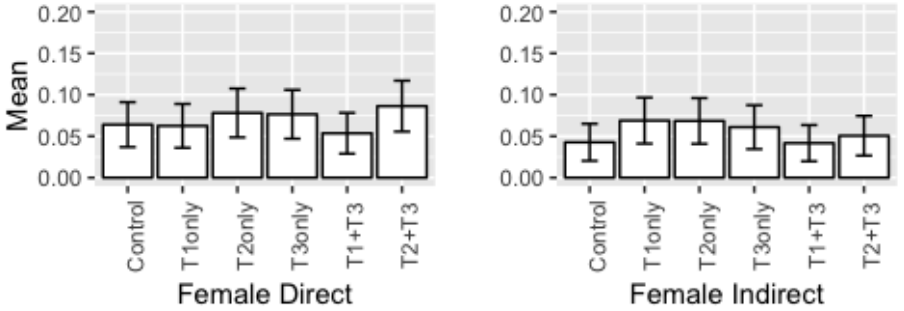


Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	-0.008	0.023*	0.001	Interaction Effect	-0.013	-0.004
	SE	0.014	0.013	0.011	SE	0.028	0.028

	Control Mean	0.1	0.083	0.096	Control Mean	0.076	0.076
					T1 or T2 Effect	-0.001	0.025*
					T3 Effect	0.007	0.007
	N	1996	1996	1996	N	1996	1996
Female Indirect	Treatment Effect	-0.002	0.011	0.026**	Interaction Effect	-0.013	-0.066**
	SE	0.013	0.014	0.012	SE	0.028	0.029
	Control Mean	0.09	0.081	0.075	Control Mean	0.055	0.055
					T1 or T2 Effect	0.004	0.044
					T3 Effect	0.052**	0.052**
	N	1998	1998	1998	N	1998	1998
Male Direct	Treatment Effect	-0.019	-0.005	0.004	Interaction Effect	-0.002	-0.054*
	SE	0.015	0.014	0.011	SE	0.03	0.028
	Control Mean	0.058	0.054	0.051	Control Mean	0.053	0.053
					T1 or T2 Effect	-0.019	0.022
					T3 Effect	0.023	0.023
	N	1496	1496	1496	N	1496	1496
Male Indirect	Treatment Effect	-0.005	0.008	-0.007	Interaction Effect	-0.015	-0.059**
	SE	0.011	0.012	0.009	SE	0.023	0.023
	Control Mean	0.056	0.051	0.055	Control Mean	0.049	0.049
					T1 or T2 Effect	0.002	0.038
					T3 Effect	0.018	0.018
	N	1987	1987	1987	N	1987	1987

RESULT 25: Prescriptive norms on public harassment perpetration (Women)

Description of measure: This measures the perceived approval of public sexual harassment perpetration within a respondent's social network of women (PNORM_PPV_F). It is coded as 1 if a respondent thinks that most out of 5 married women would approve of a man sexually harassing a woman on the street using SN8 in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	-0.021*	0.001	0.003	Interaction Effect	-0.037	0
	SE	0.012	0.012	0.01	SE	0.024	0.023
	Control Mean	0.076	0.064	0.068	Control Mean	0.064	0.064
					T1 or T2 Effect	-0.002*	0.001
					T3 Effect	0.016	0.016
	N	1996	1996	1996	N	1996	1996
Female Indirect	Treatment Effect	0.002	0.005	-0.009	Interaction Effect	-0.053**	-0.027
	SE	0.011	0.011	0.01	SE	0.024	0.022
	Control Mean	0.056	0.054	0.06	Control Mean	0.043	0.043
					T1 or T2 Effect	0.028	0.018
					T3 Effect	0.018	0.018
	N	1998	1998	1998	N	1998	1998

Male Direct	Treatment Effect	-0.004	-0.001	0.003	Interaction Effect	0.012	0.028
	SE	0.009	0.009	0.007	SE	0.02	0.018
	Control Mean	0.033	0.032	0.033	Control Mean	0.041	0.041
					T1 or T2 Effect	-0.01	-0.015
					T3 Effect	-0.011	-0.011
	N	1498	1498	1498	N	1498	1498
Male Indirect	Treatment Effect	-0.002	-0.004	-0.007	Interaction Effect	-0.007	0.004
	SE	0.008	0.008	0.007	SE	0.016	0.016
	Control Mean	0.03	0.032	0.035	Control Mean	0.037	0.037
					T1 or T2 Effect	0.002	-0.006
					T3 Effect	-0.006	-0.006
	N	1987	1987	1987	N	1987	1987

Qualitative findings on social norms around public VAWG

The qualitative findings are broadly supportive of the quantitative data, highlighting the pervasive nature of harmful norms around VAWG in public spaces.

Across the women's FGDs, women are generally intolerant of public VAWG and are very aware of the negative impact on women and girls. However, in both T1 and T2, we still see many underlying individual attitudes and norms, which blame girls and their parents for violence against them. In particular, girls can be blamed for encouraging harassment and violence because of their 'inappropriate' clothing or behaviour.

> *'When they wear such dress, they will be harassed. First thing they shouldn't have to go out in night. Such girls want that boys pass comment on them…They do not wear dupatta. No one wears dupatta nowadays. It all invites boys and men and they whistle and harass. It girls wear decent clothes, no one can dare to say anything.'* (Woman DB, T1 slum, Bhopal)

> *'Yes, sometimes girls are also wrong. Girls attract boys. Look at them, pass smile. All boys are same. Some boys just harass girl if she is going on her way. Sometimes it is the mistake of boy and sometime it is of girl. Boys sing song, chase girls. It is also harassment. Singing song, whistling is also harassment.'* (Woman DB, T2 slum, Gwalior)

T3 MDs seem more likely to recognise the negative impact of public VAWG on women and girls, in comparison to control boys who are more likely to minimise its impact. We also see some shifts in individual attitudes of men and boys blaming women and girls for violence against them because of their clothing. The idea that they should treat all women and girls like their mothers and sisters (as taught in the training they received) appears in particular to have resonated with men and boys. Some boys also report being afraid of what others in their group might think – indicating a shift in social expectations among the group. By comparison control men and boys have strong shared opinions on women being to blame for public VAWG or that men are the primary victims of harassment from women.

> *'The guys have been turned wild by the girls, they get so scantily dressed. The kind of pants men used to wear are now wore by women. Just go to the mall nearby. You will see girls dressed in jeans and guys going after them'* (Male DB control slum, Jabalpur)

> *'We should not sing songs when we see them. We should not pass comments. We should not follow them. We have sisters and mothers at our own home too.'* (Male DB, T3 slum, Bhopal)

> *'We are scared if someone sees us doing something wrong, what will they think about us.'* (Male DB, T3 slum, Bhopal)

Despite these more positive attitudes, across the intervention and control slums, participants report severe social sanctions against women survivors of public VAWG and their families– including shame, loss of reputation and honour, impact on mobility, impact on marriage prospects and even an increase in domestic violence.

On the other hand, there remain seemingly little social repercussions for men as perpetrators within the wider area, although boys harassing girls from their own slum – may face some social sanctions from the neighbourhood.

> *'People will taunt her and her parents. The family will be insulted. Whenever they will walk on roads they will insult her. She may have trouble getting married'* (Woman, T1, Jabalpur)
>
> *'Brother because of eve-teasing nothing happens. Nobody is punished. Most men and boys do it'.* (Male DB, T3, Bhopal)
>
> *Girls are even scared of revealing it [their experience of public VAW]. The family blames the girl saying she must have walked in a wrong way. They scold girls instead of scolding boys. Neighbours and the community blame the girl'* (FD, T2, Bhopal)

Reporting VAW

This sub section includes results for all of the intermediate outcome measures relating to reporting VAW including:

- Attitudes towards reporting IPV
- Attitudes towards reporting violence and harassment in public spaces
- Women reporting IPV to police or protection officer
- Women reporting public VAW to police or protection officer
- Descriptive norms around reporting IPV
- Prescriptive norms around reporting IPV
- Descriptive norms around reporting public VAW
- Prescriptive norms around reporting public VAW

Attitudes towards reporting VAW

The survey asked respondents their attitudes towards reporting IPV. While around one quarter of women report that women *should not* report husbands to the police in the case of IPV, nearly three quarters report that they *should.* For men around one third suggest they should not while around 60% suggest they should. There is weak evidence that T2 women support reporting at higher rates (75% versus 71%) but no evidence of similar treatment effects from T3 for men. (Results 26)

Individual attitudes towards reporting public VAWG are very positive with most men and women believing that women should report public VAWG to the police. There is no evidence of positive treatment effects on attitudes towards reporting public VAWG. The only evidence for effects is a worsening of attitudes, relative to programmes goals, among boys and men direct beneficiaries in T1. However, we do not find the same effect for indirect male beneficiaries. (Result 27)

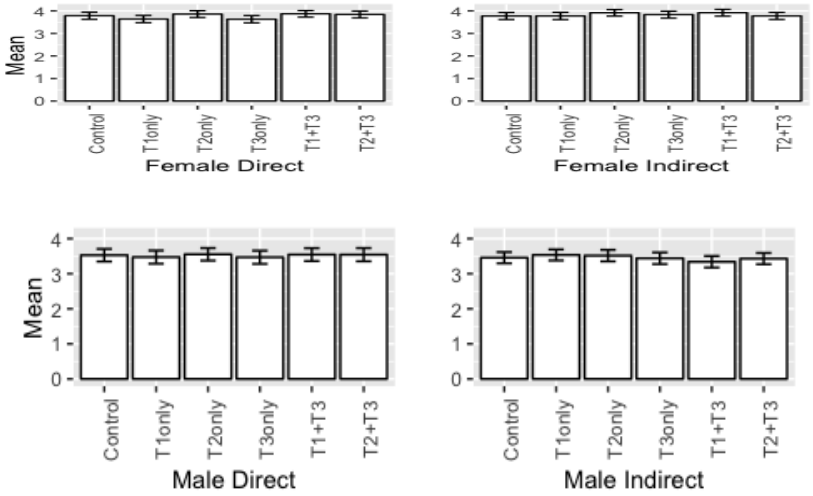
Women reporting VAWG to police or protection officer

The survey asked women how likely they would be to report an incident of IPV and violence and harassment in a public space to the police or a protection officer from a scale of 1 to 4, with 1 being very unlikely and 4 very likely. Women were more likely to answer positively to reporting an incident of violence and harassment in public than IPV. Most women report that they are somewhat likely to report both IPV and violence and harassment in public, although women are more likely to say they would report the latter (mean values of about 2.6 for IPV compared to 3.2 for violence and harassment in public). Given the very low incidence of actual reporting, this relatively high rate may suggest desirability biases in responses.

There is no evidence of positive treatment effects on reporting either IPV or public VAW. There is some limited evidence of adverse effects for FDs and FIs in T1 who are slightly less likely to say they would report IPV and violence and harassment in public spaces respectively. (Results 28 and 29)

RESULT 26: Individual attitudes on reporting IPV

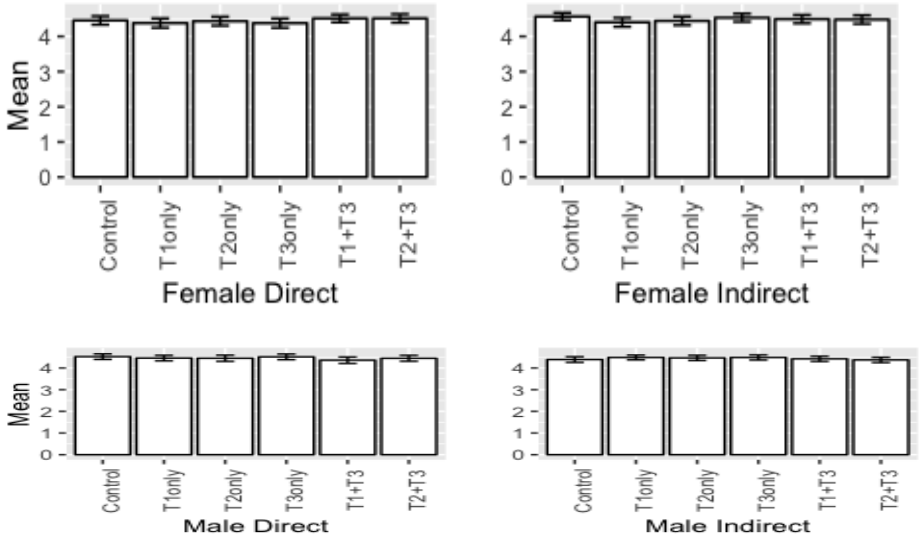
Description of measure: This measures the extent to which a respondent agrees that a woman should report her husband to the police when he hits her (ATT_RIPV). It is coded as on a scale of 1-5 where 1 is strongly disagree and 5 is strongly agree using SN17 in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	0.05	0.128*	0.034	Interaction Effect	0.467**	0.181
	SE	0.073	0.071	0.06	SE	0.144	0.142
	Control Mean	3.778	3.73	3.758	Control Mean	3.784	3.784
					T1 or T2 Effect	-0.184	0.038*
					T3 Effect	-0.184	-0.184
	N	1996	1996	1996	N	1996	1996
Female Indirect	Treatment Effect	0.038	0.035	0.034	Interaction Effect	0.114	-0.133
	SE	0.075	0.073	0.058	SE	0.148	0.142

RESULT 27: Individual attitudes on reporting public harassment

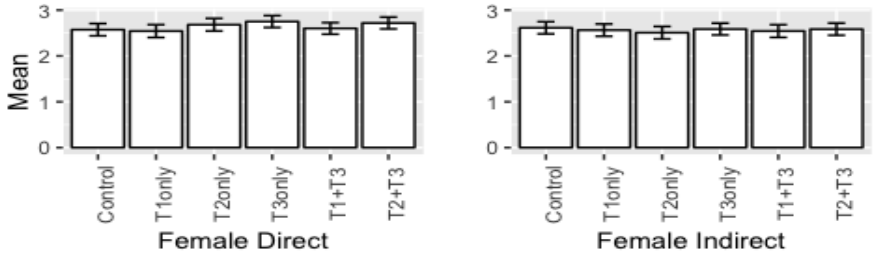
Description of measure: This measures the extent to which a respondent agrees that a woman should report to the police when she experiences sexual harassment (ATT_RPV). It is coded as on a scale of 1-5 where 1 is strongly disagree and 5 is strongly agree using SN18 in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	0.053	0.064	0.038	Interaction Effect	0.238**	0.105
	SE	0.054	0.056	0.043	SE	0.105	0.112
	Control Mean	4.444	4.43	4.421	Control Mean	4.454	4.454
					T1 or T2 Effect	-0.066	0.012
					T3 Effect	-0.077	-0.077
	N	1996	1996	1996	N	1996	1996
Female Indirect	Treatment Effect	-0.08	-0.076	0.018	Interaction Effect	0.125	-0.021
	SE	0.051	0.054	0.045	SE	0.108	0.109
	Control Mean	4.503	4.496	4.468	Control Mean	4.564	4.564
					T1 or T2 Effect	-0.143	-0.066
					T3 Effect	-0.017	-0.017
	N	1998	1998	1998	N	1998	1998
Male Direct	Treatment Effect	-0.118**	-0.07	-0.036	Interaction Effect	-0.045	-0.011
	SE	0.058	0.059	0.048	SE	0.114	0.117
	Control Mean	4.482	4.461	4.475	Control Mean	4.522	4.522
					T1 or T2 Effect	-0.096**	-0.064
					T3 Effect	-0.017	-0.017
	N	1498	1498	1498	N	1498	1498
Male Indirect	Treatment Effect	0.008	-0.01	-0.021	Interaction Effect	-0.115	-0.218**
	SE	0.049	0.054	0.041	SE	0.095	0.101
	Control Mean	4.427	4.445	4.446	Control Mean	4.388	4.388
					T1 or T2 Effect	0.066	0.1
					T3 Effect	0.091	0.091
	N	1988	1988	1988	N	1988	1988

RESULT 28: Reporting IPV to the police or a protection officer

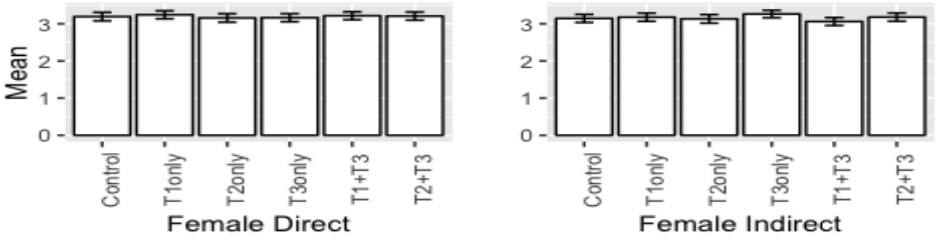
Description of measure: This measures the likelihood of reporting a hypothetical incident of hitting or choking by an intimate partner to the police, among married/cohabiting women (RIPV_POLICE). It is coded as on a scale of 1-4, where 1 is very unlikely to report and 4 is very likely to report using RDV9 in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	-0.122**	0.009	0.093*	Interaction Effect	-0.026	-0.118
	SE	0.059	0.056	0.049	SE	0.118	0.111
	Control Mean	2.681	2.615	2.599	Control Mean	2.573	2.573
					T1 or T2 Effect	-0.108**	0.068
					T3 Effect	0.142*	0.142*
	N	1754	1754	1754	N	1754	1754
Female Indirect	Treatment Effect	-0.066	-0.077	0.02	Interaction Effect	0.135	0.131
	SE	0.063	0.063	0.051	SE	0.124	0.13
	Control Mean	2.573	2.577	2.561	Control Mean	2.616	2.616
					T1 or T2 Effect	-0.135	-0.143
					T3 Effect	-0.07	-0.07
	N	1660	1660	1660	N	1660	1660

RESULT 29: Reporting public violence or harassment to a police or protection officer

Description of measure: This measures the likelihood of reporting a hypothetical incident of stalking or groping in public place to the police, among women (RPV_POLICE). It is coded as on a scale of 1-5, where 1 is very unlikely to report and 5 is very likely to report using RPV1 in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	0.047	-0.012	-0.03	Interaction Effect	0.073	-0.017
	SE	0.05	0.053	0.041	SE	0.098	0.104
	Control Mean	3.179	3.203	3.197	Control Mean	3.192	3.192
					T1 or T2 Effect	0.011	-0.004
					T3 Effect	-0.049	-0.049
	N	1996	1996	1996	N	1996	1996
Female Indirect	Treatment Effect	-0.095**	-0.071	0.007	Interaction Effect	-0.197**	-0.137
	SE	0.047	0.048	0.04	SE	0.095	0.098
	Control Mean	3.178	3.16	3.15	Control Mean	3.143	3.143
					T1 or T2 Effect	0.004**	-0.003
					T3 Effect	0.119	0.119
	N	1998	1998	1998	N	1998	1998

Social norms around reporting IPV and public VAWG

The survey also sought to measure social norms around reporting IPV and public VAWG – including both descriptive and prescriptive norms.

Respondents expect that only 20% of women would report IPV to the police, which is in sharp contrast to the prevailing individual attitude that IPV *should* be reported to the police. This finding suggests a disconnect between descriptive norms and individual attitudes towards reporting IPV, resulting either from false beliefs about how others would act or, perhaps, a failure to report IPV even when women individually think that they should. We see no evidence that the different treatments had any effect on descriptive norms around reporting IPV. (Result 30)

About 54% of women and 40% of men believe that most married men in their social network would approve of a woman reporting IPV to the police. Both men and women expect women to be more approving of reporting IPV to the police than men. About 70% of women and 52% of men believe that women would approve of a woman who reported IPV. This finding may indicate that prescriptive norms among men and boys present a bigger barrier to reporting IPV than among women, highlighting the importance of engaging men and boys on the issue of VAW. (Result 32 and 33)

We see very little evidence of treatment effects on prescriptive norms around reporting, with the exception that men in T3 are more likely to expect that most women in their social network will be more likely to approve of other women reporting IPV to the police. However, we do not see the same effect on their expectation of men in their social network. (Result 32 and 33)

Women and men expect that women would be more likely to report an incident of violence and harassment in public to the police than IPV. Respondents believe that roughly half of the women in their social network would be likely to report violence and harassment in public, with men believing that women would report at slightly higher rates (60% compared to 50%). (Result 31)

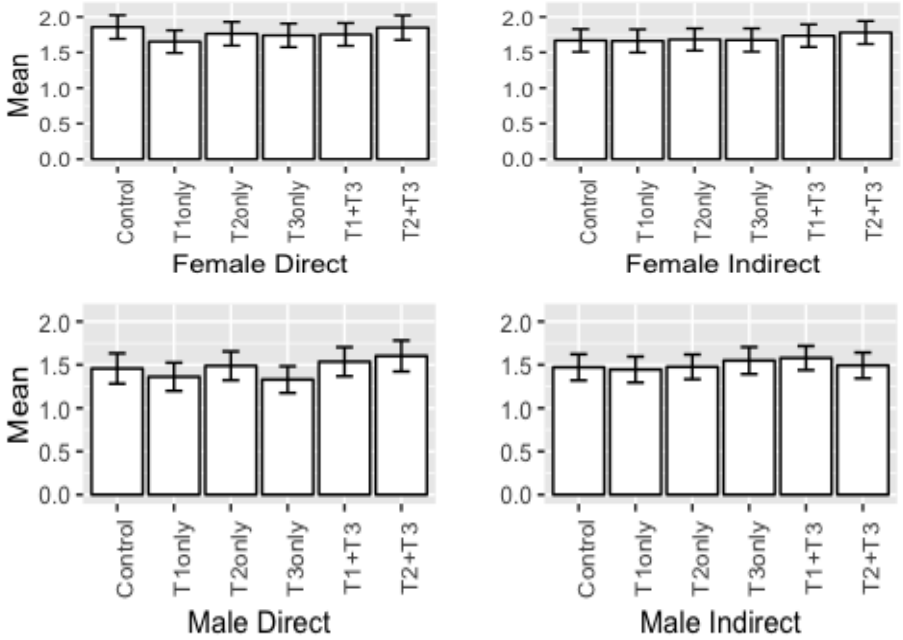
We see limited and inconsistent evidence of treatment effects on descriptive norms of reporting public VAWG. We find a weak and small effect on potential MDs in the T2 intervention with slightly more men in this group believing that most women would report public VAWG to the police. However, we see the opposite effect on MIs for both the T1 and T2 groups. (Result 31)

Women and men's expectations of social approval of reporting are much higher than expectations of how many women would report, with women and men agreeing that most men in their social network would approved of a woman reporting IPV – slightly higher for women. (Result 34 and 35)

Again we see limited and inconsistent evidence of treatment effects on prescriptive norms of reporting public VAWG. We see a strong positive effect on MDs in the T3 intervention with men believing slightly more women would approve of reporting public VAWG. However, there is a weak negative effect on the same group in the T2 intervention. (Result 35). There is also a small negative effect on MIs expectations of men's social approval of reporting public VAWG from the T1 and T2 intervention – however we do not see the same effect on MDs. (Result 34)

RESULT 30: Descriptive norms on reporting IPV

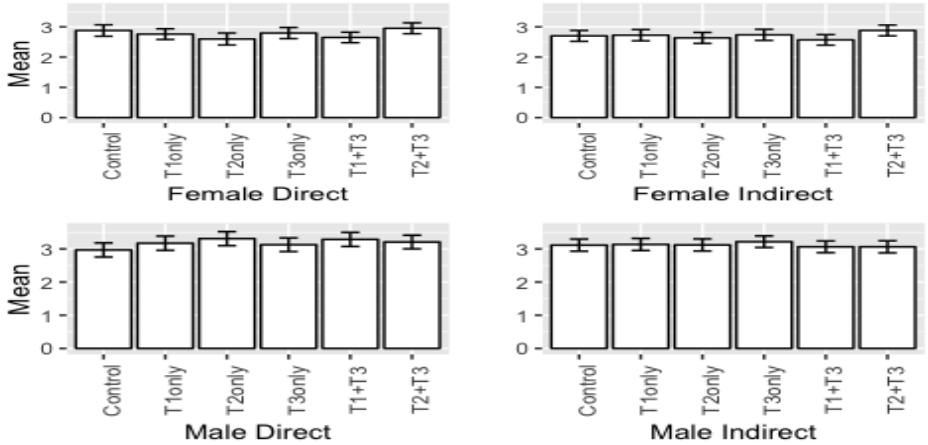
Description of measure: This measures the perceived prevalence of IPV reporting within a respondent's social network (DNORM_RIPV). It is coded as the number (out of 5 married women) that a respondent believes would report to the police if their husband hit them using SN9 in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	-0.091	0.024	0.018	Interaction Effect	0.208	0.202
	SE	0.073	0.071	0.06	SE	0.145	0.141
	Control Mean	1.802	1.75	1.756	Control Mean	1.857	1.857
					T1 or T2 Effect	-0.195	-0.077
					T3 Effect	-0.119	-0.119
	N	1996	1996	1996	N	1996	1996
Female Indirect	Treatment Effect	-0.005	0.033	0.074	Interaction Effect	0.033	0.138
	SE	0.079	0.076	0.062	SE	0.162	0.156
	Control Mean	1.701	1.685	1.67	Control Mean	1.668	1.668
					T1 or T2 Effect	-0.021	-0.036
					T3 Effect	0.017	0.017
	N	1998	1998	1998	N	1998	1998
Male Direct	Treatment Effect	0.03	0.125*	0.071	Interaction Effect	0.31**	0.306**
	SE	0.073	0.073	0.06	SE	0.149	0.145
	Control Mean	1.47	1.421	1.435	Control Mean	1.457	1.457
					T1 or T2 Effect	-0.125	-0.028*
					T3 Effect	-0.136	-0.136
	N	1498	1498	1498	N	1498	1498
Male Indirect	Treatment Effect	-0.032	-0.038	0.078	Interaction Effect	0.043	-0.034
	SE	0.076	0.07	0.058	SE	0.153	0.139
	Control Mean	1.497	1.511	1.464	Control Mean	1.471	1.471
					T1 or T2 Effect	-0.054	-0.021
					T3 Effect	0.075	0.075
	N	1988	1988	1988	N	1988	1988

RESULT 31: Descriptive norms on reporting Public VAWG

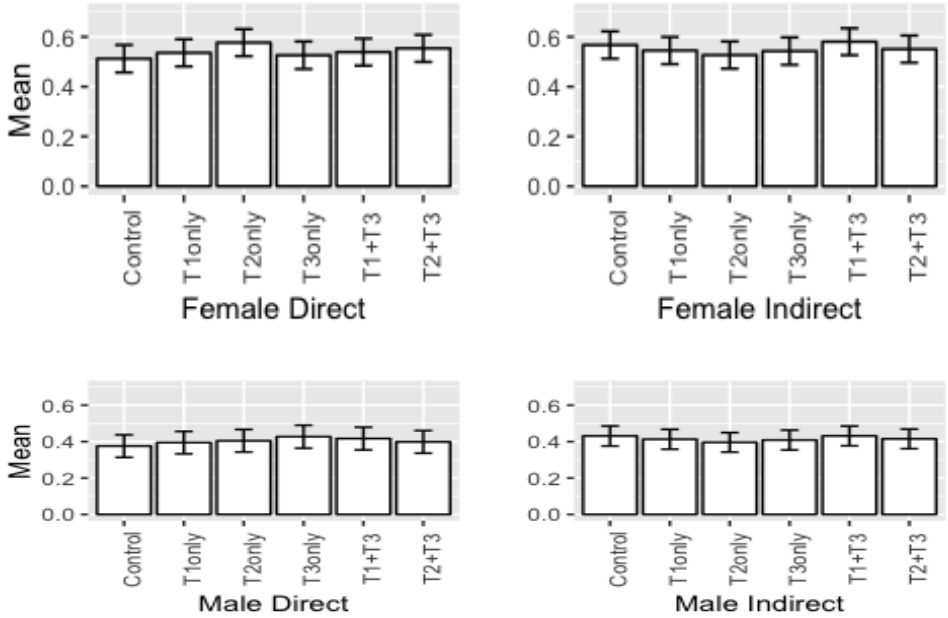
Description of measure: This measures the perceived prevalence of public sexual harassment reporting within a respondent's social network (DNORM_RPV). It is coded as the number (out of 5 married women) that a respondent believes would report to the police if they experienced public sexual harassment using SN12 in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	-0.077	-0.018	0.048	Interaction Effect	0.15	0.332**
	SE	0.077	0.076	0.059	SE	0.153	0.145
	Control Mean	2.806	2.769	2.744	Control Mean	2.878	2.878
					T1 or T2 Effect	-0.152	-0.184
					T3 Effect	-0.114	-0.114
	N	1995	1995	1995	N	1995	1995
Female Indirect	Treatment Effect	-0.025	0.064	0.04	Interaction Effect	-0.062	0.091
	SE	0.074	0.079	0.065	SE	0.155	0.159
	Control Mean	2.738	2.682	2.686	Control Mean	2.701	2.701
					T1 or T2 Effect	0.006	0.019
					T3 Effect	0.03	0.03
	N	1998	1998	1998	N	1998	1998
Male Direct	Treatment Effect	0.149	0.175*	0.077	Interaction Effect	0.131	-0.142
	SE	0.093	0.091	0.073	SE	0.18	0.18
	Control Mean	3.154	3.139	3.151	Control Mean	2.967	2.967
					T1 or T2 Effect	0.083	0.247*
					T3 Effect	0.081	0.081
	N	1497	1497	1497	N	1497	1497
Male Indirect	Treatment Effect	-0.154**	-0.158**	0.017	Interaction Effect	0.012	-0.164
	SE	0.073	0.076	0.057	SE	0.143	0.154
	Control Mean	3.128	3.131	3.122	Control Mean	3.113	3.113
					T1 or T2 Effect	-0.159**	-0.075**
					T3 Effect	0.069	0.069
	N	1987	1987	1987	N	1987	1987

RESULT 32: Prescriptive norms on reporting IPV (Men)

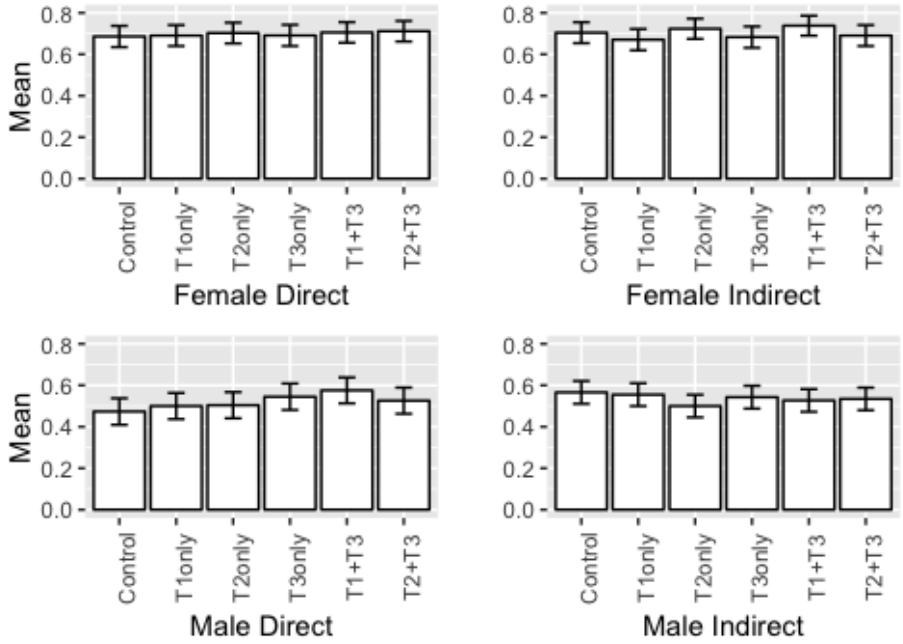
Description of measure: This measures the perceived approval of IPV reporting within a respondent's social network of men (PNORM_RIPV_M). It is coded as 1 if a respondent thinks that most out of 5 married men would approve of a woman reporting her husband to the police if he hits her using SN10A in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	0.013	0.044	-0.004	Interaction Effect	-0.045	-0.045
	SE	0.029	0.029	0.023	SE	0.057	0.056
	Control Mean	0.542	0.528	0.542	Control Mean	0.512	0.512
					T1 or T2 Effect	0.036	0.066
					T3 Effect	0.026	0.026
	N	1996	1996	1996	N	1996	1996
Female Indirect	Treatment Effect	0	-0.02	0.016	Interaction Effect	0.025	0.05
	SE	0.025	0.026	0.021	SE	0.053	0.054
	Control Mean	0.547	0.559	0.546	Control Mean	0.567	0.567
					T1 or T2 Effect	-0.013	-0.045
					T3 Effect	-0.009	-0.009
	N	1998	1998	1998	N	1998	1998
Male Direct	Treatment Effect	-0.005	0	0.024	Interaction Effect	0.008	-0.031
	SE	0.026	0.027	0.021	SE	0.052	0.052
	Control Mean	0.401	0.403	0.392	Control Mean	0.376	0.376
					T1 or T2 Effect	-0.009	0.016
					T3 Effect	0.032	0.032
	N	1497	1497	1497	N	1497	1497
Male Indirect	Treatment Effect	0	-0.01	0	Interaction Effect	0.044	0.035
	SE	0.022	0.024	0.019	SE	0.044	0.049
	Control Mean	0.412	0.421	0.413	Control Mean	0.431	0.431
					T1 or T2 Effect	-0.022	-0.028
					T3 Effect	-0.026	-0.026
	N	1985	1985	1985	N	1985	1985

RESULT 33: Prescriptive norms on reporting IPV (Women)

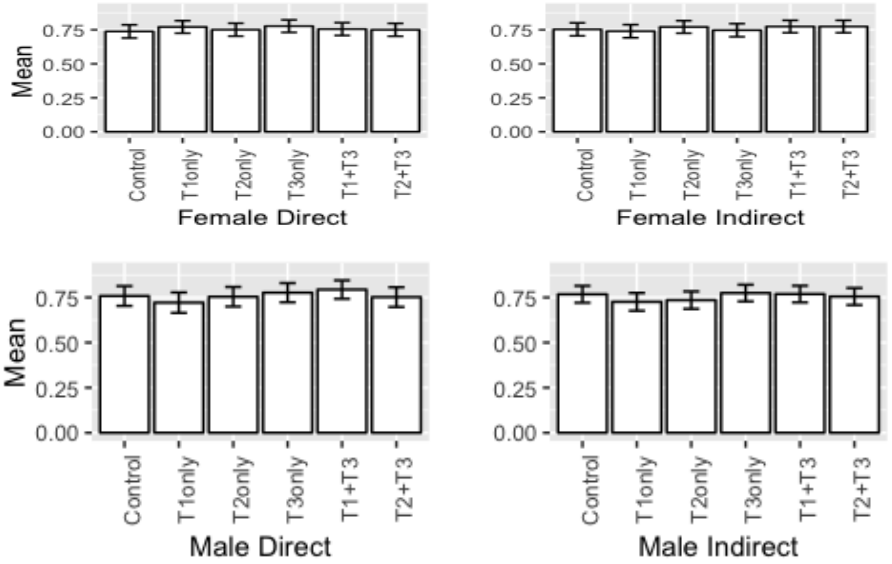
Description of measure: This measures the perceived approval of IPV reporting within a respondent's social network of women (PNORM_RIPV_F). It is coded as 1 if a respondent thinks that most out of 5 married women would approve of a woman reporting her husband to the police if he hits her using SN11A in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	0.013	0.034	0.004	Interaction Effect	-0.024	-0.009
	SE	0.026	0.022	0.02	SE	0.052	0.043
	Control Mean	0.698	0.693	0.693	Control Mean	0.686	0.686
					T1 or T2 Effect	0.026	0.038
					T3 Effect	0.015	0.015
	N	1996	1996	1996	N	1996	1996
Female Indirect	Treatment Effect	0.001	0.016	0.006	Interaction Effect	0.081*	0.006
	SE	0.023	0.023	0.018	SE	0.045	0.045
	Control Mean	0.7	0.699	0.699	Control Mean	0.704	0.704
					T1 or T2 Effect	-0.039	0.013
					T3 Effect	-0.024	-0.024
	N	1998	1998	1998	N	1998	1998
Male Direct	Treatment Effect	0.017	0.007	0.054**	Interaction Effect	0.044	-0.029
	SE	0.03	0.032	0.024	SE	0.059	0.062
	Control Mean	0.512	0.524	0.493	Control Mean	0.473	0.473
					T1 or T2 Effect	-0.005	0.021
					T3 Effect	0.049**	0.049**
	N	1498	1498	1498	N	1498	1498
Male Indirect	Treatment Effect	-0.027	-0.035	-0.006	Interaction Effect	0.034	0.078*
	SE	0.022	0.022	0.018	SE	0.044	0.044
	Control Mean	0.535	0.547	0.54	Control Mean	0.566	0.566
					T1 or T2 Effect	-0.044	-0.075
					T3 Effect	-0.044	-0.044
	N	1988	1988	1988	N	1988	1988

RESULT 34: Prescriptive norms on reporting public harassment (Men)

Description of measure: This measures the perceived approval of public sexual harassment reporting within a respondent's social network of men (PNORM_RPV_M). It is coded as 1 if a respondent thinks that most of out of 5 married men would approve of a woman reporting to the police if she experienced public sexual harassment using SN13 in the endline survey instrument.

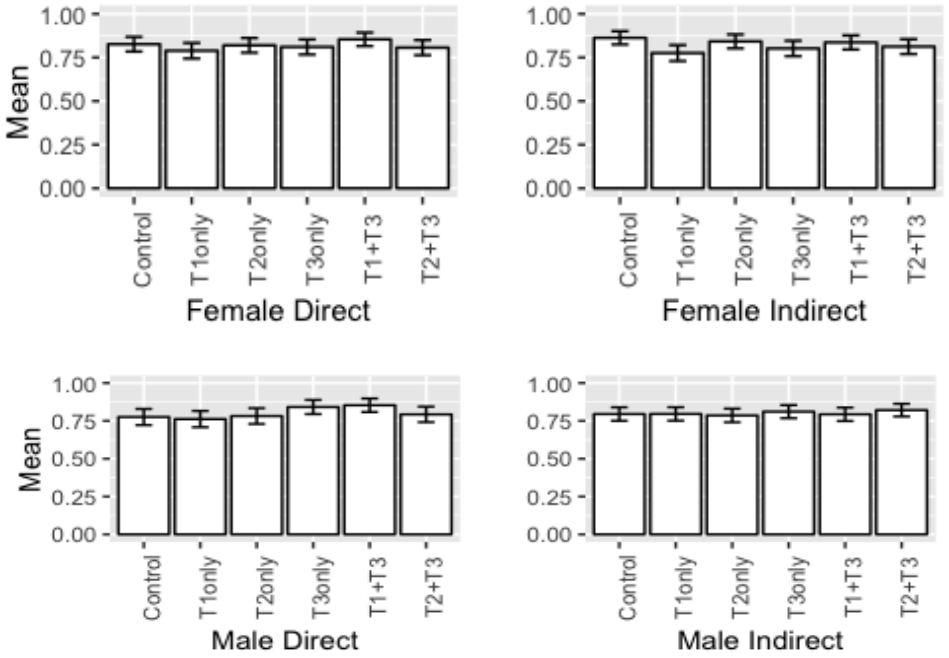


Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	0.011	0.005	0.005	Interaction Effect	-0.046	-0.056
	SE	0.025	0.022	0.02	SE	0.048	0.045
	Control Mean	0.754	0.76	0.753	Control Mean	0.738	0.738
					T1 or T2 Effect	0.034	0.034
					T3 Effect	0.039	0.039
	N	1995	1995	1995	N	1995	1995
Female Indirect	Treatment Effect	0.001	0.026	0.011	Interaction Effect	0.043	0
	SE	0.021	0.02	0.017	SE	0.041	0.041
	Control Mean	0.761	0.753	0.755	Control Mean	0.753	0.753
					T1 or T2 Effect	-0.02	0.026
					T3 Effect	-0.003	-0.003
	N	1998	1998	1998	N	1998	1998
Male Direct	Treatment Effect	-0.036	-0.036	0.035*	Interaction Effect	0.084*	-0.003
	SE	0.023	0.023	0.019	SE	0.045	0.046
	Control Mean	0.76	0.762	0.744	Control Mean	0.758	0.758
					T1 or T2 Effect	-0.079	-0.034
					T3 Effect	0.008*	0.008*
	N	1495	1495	1495	N	1495	1495
Male Indirect	Treatment Effect	-0.044**	-0.04**	0.03*	Interaction Effect	0.057	0.035
	SE	0.019	0.019	0.016	SE	0.038	0.038
	Control Mean	0.758	0.759	0.743	Control Mean	0.768	0.768
					T1 or T2 Effect	-0.073**	-0.057**
					T3 Effect	-0.001*	-0.001*
	N	1987	1987	1987	N	1987	1987

See above.

RESULT 35: Prescriptive norms on reporting public harassment (Women)

Description of measure: This measures the perceived approval of public sexual harassment reporting within a respondent's social network of women (PNORM_RPV_F). It is coded as 1 if a respondent thinks that most of out of 5 married women would approve of a woman reporting to the police if she experienced public sexual harassment using SN14 in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	0.01	0.011	0.011	Interaction Effect	0.105**	-0.01
	SE	0.018	0.02	0.016	SE	0.035	0.039
	Control Mean	0.816	0.82	0.811	Control Mean	0.826	0.826
					T1 or T2 Effect	-0.043	0.016
					T3 Effect	-0.021	-0.021
	N	1996	1996	1996	N	1996	1996
Female Indirect	Treatment Effect	-0.029	0.001	-0.01	Interaction Effect	0.13**	0.021
	SE	0.02	0.019	0.016	SE	0.041	0.037
	Control Mean	0.83	0.819	0.827	Control Mean	0.863	0.863
					T1 or T2 Effect	-0.094	-0.009
					T3 Effect	-0.06	-0.06
	N	1998	1998	1998	N	1998	1998
Male Direct	Treatment Effect	-0.019	-0.041*	0.052**	Interaction Effect	0.054	-0.045
	SE	0.024	0.021	0.019	SE	0.045	0.043
	Control Mean	0.798	0.808	0.773	Control Mean	0.776	0.776
					T1 or T2 Effect	-0.046	-0.019*
					T3 Effect	0.049**	0.049**
	N	1498	1498	1498	N	1498	1498
Male Indirect	Treatment Effect	-0.021	-0.009	0.014	Interaction Effect	0.023	0.03
	SE	0.018	0.019	0.014	SE	0.036	0.037
	Control Mean	0.803	0.799	0.792	Control Mean	0.795	0.795
					T1 or T2 Effect	-0.033	-0.024
					T3 Effect	-0.003	-0.003
	N	1988	1988	1988	N	1988	1988

Qualitative findings on reporting VAWG

Awareness of reporting and response mechanisms is also higher among SHG members who have received the VAWG training (T2). In T1 and control slums, there is low awareness of reporting mechanisms, other than the police. In T2 slums, FDs are aware of specific forms of IPV and public VAWG which are against the law, and multiple reporting mechanisms (including police, women's cell, protection officer, Aaganvari centre, Nirbhhaya teams and the domestic violence helpline).

> *'If she is aware of us, she can come to us and ask for help. We will ask her to go to the protection officer, aanganvari or child and women development officer.'* (Woman DB, T2 slum, Bhopal)

> *'We should go to the protection officer and complain about our husbands. He will come to make our husbands understand and will also help in all possible ways.'* (Woman DB, T2 slum, Indore)

Box 5: Qualitative exercises on reporting VAWG

The FGDs also explored norms around reporting IPV and public VAWG, and awareness of formal/informal response mechanisms including laws through a number of exercises and questions. In particular, respondents initial reactions to the video vignettes and subsequent discussions, provided rich information on social expectations around reporting and responding to IPV and public VAWG (including descriptive and injunctive norms, and social sanctions), as well as awareness, knowledge and trust in formal and informal response mechanisms.

We see little evidence of difference between the control and intervention men and boys groups in terms of awareness of laws and reporting mechanisms. In general, MDs in T3 are aware that there is a law against domestic violence but their awareness of reporting and response mechanisms is limited to the police and helpline numbers.

In T1 and control slums, reporting IPV is perceived to be extremely rare among FDs, and prescriptive norms dictate that IPV should be reported to the police only in very severe cases. In such extreme cases, reporting is seen only as a way for the wife to get a few days respite from violence – rather than as a long- term solution or as a matter of social justice. In T2 slums, reporting IPV to the police is perceived to be slightly more common and slightly more acceptable among FDs, with a women giving examples from their families and neighbours reporting instances of IPV, or using the threat of reporting IPV as a deterrent.

> *'We had gone to the sanitation department of the municipality. After returning home, one of our women was beaten up by her husband for going outside. So we went there and warned him – next time if you do such things, we will go to the police station. It was effective'.* (Woman DB, T2 slum, Indore)

However, across treatment and control slums, women fear severe repercussions for reporting IPV including increased risk of violence (perception that domestic violence will increase if husband is reported to police); as well as economic consequences (including cost to bail husband out of jail, loss of husband's earnings, and severe economic consequences for women and their children if the marriage breaks down). In addition, there are perceived severe social sanctions for women who report IPV including family breakdown, divorce, homelessness, loss of respect and honour, and the risk that the woman ostracised from their family and community.

Norms against reporting IPV are held in place by the strong and persistent harmful social norm that IPV as a family matter. This norm can also be seen to influence the logic of local response mechanisms – with reports of police emphasising family reconciliation and mediation over safety of survivor and justice. Moreover, it influences women's expectations of reporting mechanisms' primary function to help *mediate* in IPV cases - to help 'solve' domestic violence cases – rather than as a means of social justice.

> *'There are ways but you can't keep your husband arrested for long. Family members will think, the wife is bad, she has left her husband in jail'.* (Woman DB, T1 slum, Bhopal)

'They are scared of being defamed in the society. Even if they come and lodge case, later they come and tell us that they don't want to take any action'. (KII, Police Inspector, Bhopal)

Similar to findings on reporting IPV, reporting public VAWG is also perceived as uncommon due to severe social sanctions against survivors including affecting marriage prospects, loss of honour for family, and increased risk of violence – even death as retribution by perpetrators and their families. Whilst it is more acceptable to report to another family member, or if the perpetrator is from within the slum, or to try to 'solve' the issue through community channels- there are still expected to be restrictions as a result – in particular on the woman or girls mobility and the family's honour.

In T2 slums, FDs report being more supportive of reporting public VAWG than FDs in control or T1 slums. However, despite increased social support for reporting among SHG members, fear of severe social sanctions from the wider community make it very hard for women and girls to report in reality. In particular, women feel that it is particularly difficult for young girls to report instances of public VAWG – because of the likely negative response of friends and family.

Reporting VAWG is severely undermined by low trust in police response and the perceived (and in some cases experienced) inadequacies in local response mechanisms - which present significant barriers to reporting IPV and public VAWG. In particular, there is an expectation that the police will only hold the perpetrator for a few days and/or beat the husband up as a punishment, police corruption, influential and powerful elite will not be prosecuted, but that a jail term is unlikely, with the norm that the wife will ultimately have to bail their husbands out of jail.

Women's economic empowerment

This sub section includes results for all of the intermediate outcome measures relating to women's economic empowerment including:

- Women earning their own income
- Women's control of household income
- Women's role in household decision making

Women's income and control over income

About 40% of FDs, and 30% of FIs report earning their own income. We expected to see the biggest effects on earnings for the FDs, particularly in T1 and T2. In both cases however we see no evidence of treatment effects. The T1 women respondents report slightly lower rates of earnings compared to control and the T2 women slightly larger rates, however neither difference is statistically significant. There is no evidence of an effect of T3 on earning, as is to be expected.

The survey asked married women if they were a primary decision maker with regard to two types of economic decisions: major purchases and daily household needs. Most women report being the primary decision maker with regard to one of these two economic decisions.

There is no evidence of positive treatment effects on women's control over household income, and some evidence of adverse effects for FDs in T1 and T3. The drop for T1 beneficiaries is equivalent to about one in ten more women reporting to not be the primary decision maker over one of these two decisions compared to control. This adverse effect could possibly be attributed to backlash (husbands tightening control) following attempts to strengthen women's economic situation.

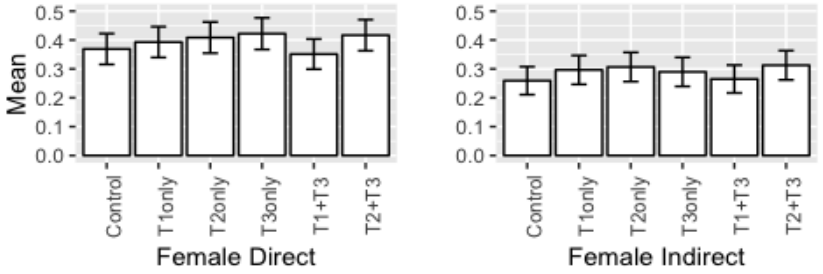
Household decision-making

Women were asked about decision-making around major household purchases, daily expenses, health care, and visit to family or neighbours. Women typically report having a primary or shared role in decision making in three out of four of these areas, with somewhat higher levels of decision-making control reported for the FDs than FIs.

There is no evidence of treatment effects on household decision-making associated with T1 or T2. There is evidence of a small adverse effect for FDs in T3 (significant at the 10% level only).

RESULT 36: Women earning their own income

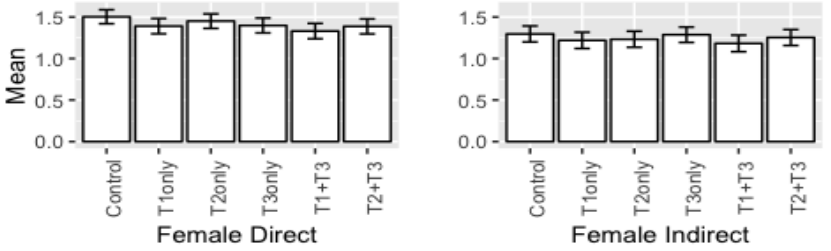
Description of measure: This measures whether women are currently working for pay (INCOME_EARN). It is coded as 1 if the respondent is currently working for pay (either in cash or in kind) and 0 if the respondent is not working or working without pay using YN1 in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	-0.01	0.034	0.004	Interaction Effect	-0.11**	-0.114**
	SE	0.026	0.026	0.021	SE	0.051	0.054
	Control Mean	0.404	0.384	0.39	Control Mean	0.369	0.369
					T1 or T2 Effect	0.045	0.091
					T3 Effect	0.079	0.079
	N	1996	1996	1996	N	1996	1996
Female Indirect	Treatment Effect	0.004	0.027	-0.001	Interaction Effect	-0.095**	-0.053
	SE	0.024	0.023	0.019	SE	0.047	0.045
	Control Mean	0.292	0.278	0.288	Control Mean	0.259	0.259
					T1 or T2 Effect	0.051	0.054
					T3 Effect	0.049	0.049
	N	1998	1998	1998	N	1998	1998

RESULT 37: Control exercised by women over household income

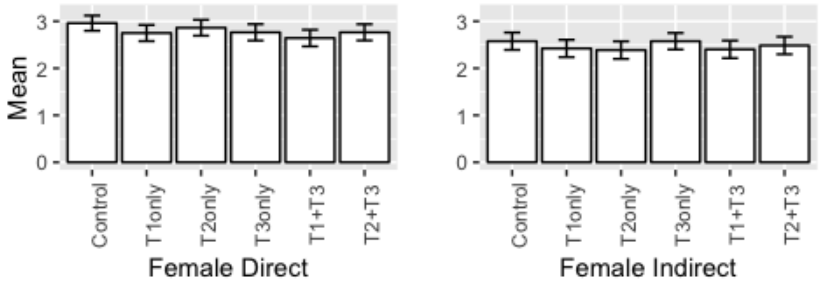
Description of the measure: This measures the control exercised by married or cohabiting women over household income (HHINCOME_CONTROL). It is coded as the number of household spending decisions, from a list of two, over which married women have primary or joint decision-making power using EM1A - B in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	-0.091**	-0.024	-0.068*	Interaction Effect	0.043	0.016
	SE	0.046	0.042	0.035	SE	0.091	0.084
	Control Mean	1.41	1.391	1.422	Control Mean	1.483	1.483
					T1 or T2 Effect	-0.113**	-0.032
					T3 Effect	-0.088*	-0.088*
	N	1754	1754	1754	N	1754	1754
Female Indirect	Treatment Effect	-0.06	-0.004	-0.03	Interaction Effect	-0.019	0.053
	SE	0.051	0.051	0.043	SE	0.104	0.1
	Control Mean	1.299	1.282	1.282	Control Mean	1.354	1.354
					T1 or T2 Effect	-0.051	-0.031
					T3 Effect	-0.042	-0.042
	N	1660	1660	1660	N	1660	1660

RESULT 38: Role of women in household decision-making

Description of measure: This measures the role played by married or cohabitating women in household decision-making (DECISIONS_ROLE). It is coded as the number of household decisions, from a list of 4, over which married women have primary or joint decision-making power using EM1A - D in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	-0.14	0.008	-0.123*	Interaction Effect	0.131	0.063
	SE	0.089	0.084	0.068	SE	0.176	0.165
	Control Mean	2.782	2.743	2.803	Control Mean	2.917	2.917
					T1 or T2 Effect	-0.206	-0.024
					T3 Effect	-0.188*	-0.188*
	N	1754	1754	1754	N	1754	1754
Female Indirect	Treatment Effect	-0.071	-0.017	-0.009	Interaction Effect	0.002	0.164
	SE	0.096	0.096	0.081	SE	0.201	0.19
	Control Mean	2.554	2.554	2.512	Control Mean	2.669	2.669
					T1 or T2 Effect	-0.072	-0.1
					T3 Effect	-0.066	-0.066
	N	1660	1660	1660	N	1660	1660

Qualitative findings on women's economic empowerment

Box 6: Qualitative exercises on women's economic empowerment

In FGDs, women were asked to reflect on most significant changes in their individual lives, their relationships, their SHG, and the wider community as a result of the intervention. These discussions touched on a number of different aspects of economic and social empowerment – which were coded and included in analysis.

The qualitative component allowed us to explore broader components of economic empowerment than the quantitative outcome measures which focus on improved income and control over income.

The qualitative data is broadly supportive of the quantitative findings showing no change in women's income or control over income. FDs feel that increased savings and access to loans are among the most significant positive changes in their lives as a result of SHG membership – providing them with the means to help solve household problems and pay for emergency expenses. However, despite these perceived benefits, there is little evidence that this has led to improved livelihoods or financial independence of women. Moreover, women identified women's unemployment and lack of viable employment opportunities as the most significant challenge they continue to face.

Consequently, any improvements in savings and access to loans (whilst appreciated by SHG members) have not on the whole led to increased financial independence. Husbands remain the main breadwinners, and their permission is often necessary for women to access loans (as well as to attend the SHG).

There were a few exceptions to these general findings. In a few slums in Jabalpur, some SHG members had managed to improve their livelihoods by investing in small income earning activities as a result of SHG activities, and these women reported greater financial independence.

The SHG intervention appears to have had an impact on broader areas of women's empowerment which may reduce women's vulnerability to VAWG in the longer term including increased self-confidence and improved social networks.

FDs in T1 and T2, report to have gained confidence, self-belief and courage through being part of a group and learning new information and skills. Women often link increased confidence with improved mobility reporting to be less fearful to go out and interact with others.

Improved social networks are also regarded by women as one of the most significant improvements in their lives from the intervention. There is wider international evidence that social networks reduce women's vulnerability to VAWG[60]. FD report that their friendships have improved and that they feel more confident being part of the group. Furthermore, FDs report to have gained recognition and respect from other members in their community for their work – which has further boosted their confidence and feelings of self-worth.

Women's mobility and feelings of safety

This sub section includes results for all of the intermediate outcome measures relating to women's mobility and feelings of safety including:

- Women's mobility outside their slum
- Women's feelings of safety in public spaces during the day
- Women's feelings of safety in public spaces at night

Women's mobility

Women were asked how often they travelled outside their home slum in the past three months. FDs report travelling outside of their slums at slightly higher rates than FIs with women on average reporting to travel outside a few times a month.

There is no evidence that FDs have increased mobility as a result of T1 or T2. However, there is some evidence that FIs have greater mobility following T1 and T2. It is difficult to account for effects on FIs, without direct effects on FDs. As discussed in the discussion below, there is no evidence that any of the increased mobility is associated is due to increased sense of safety. (Result 39)

Women's feelings of safety in public spaces

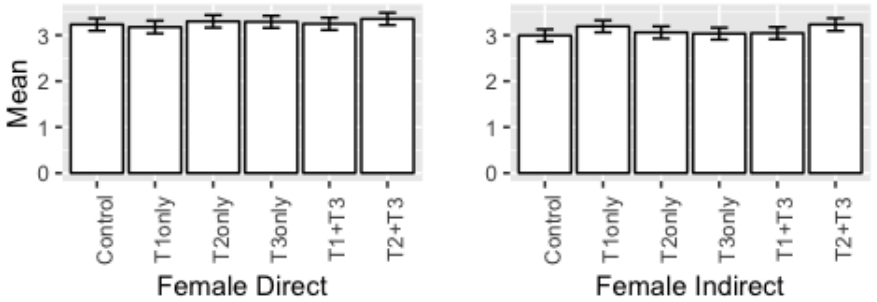
Women were asked how safe they feel in and around their home slum during both the day and night. Women report to feel safe visiting most public sites (8 out of 11) during the day time, but less than half of these sites (4 out of 11) in the night time. Reports of safety were nearly identical across all treatment combinations.

There is no evidence of treatment effects on feelings of safety in public spaces either during the day or at night. (Result 40 and 41)

[60] Population council (2005)

RESULT 39: Mobility of women outside their slum

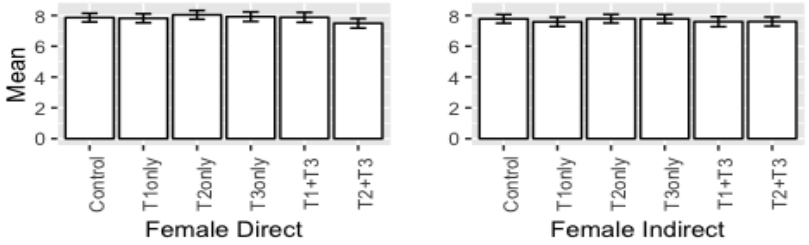
Description of measure: This measures the mobility of women outside the colony during the day (MOBILITY). It is coded as the frequency with which women travelled outside their home slum in the past 3 months on a scale from 0 (Never) to 5 (Everyday/Almost Everyday) using MS1 in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	-0.041	0.084	0.075	Interaction Effect	0.027	-0.085
	SE	0.059	0.062	0.051	SE	0.123	0.127
	Control Mean	3.295	3.236	3.236	Control Mean	3.232	3.232
					T1 or T2 Effect	-0.054	0.126
					T3 Effect	0.095	0.095
	N	1996	1996	1996	N	1996	1996
Female Indirect	Treatment Effect	0.108*	0.127**	0.04	Interaction Effect	-0.25**	0.053
	SE	0.062	0.063	0.05	SE	0.126	0.128
	Control Mean	3.081	3.066	3.082	Control Mean	2.994	2.994
					T1 or T2 Effect	0.233*	0.1**
					T3 Effect	0.106	0.106
	N	1998	1998	1998	N	1998	1998

RESULT 40: Feelings of safety in public spaces during the day among women

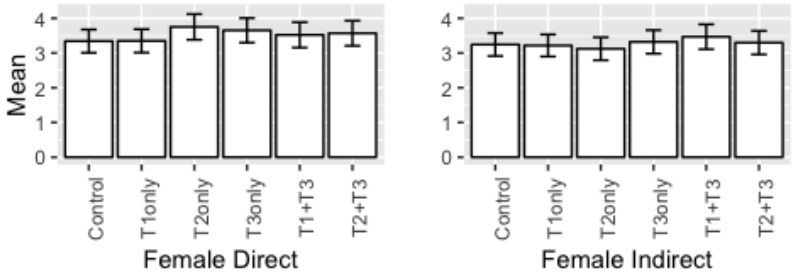
Description of measure: This measures the feelings of safety in public spaces in and around the home slum during the day among women (SAFETY_DAY). It is coded as the number of public places, from a list of 11, that women feel safe or very safe visiting alone during the day using MS3 A-K in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	-0.089	-0.176	-0.139	Interaction Effect	0.086	-0.564
	SE	0.196	0.18	0.149	SE	0.387	0.362
	Control Mean	7.823	7.865	7.902	Control Mean	7.857	7.857
					T1 or T2 Effect	-0.132	0.106
					T3 Effect	0.021	0.021
	N	1996	1996	1996	N	1996	1996
Female Indirect	Treatment Effect	-0.239	-0.141	-0.012	Interaction Effect	-0.019	-0.284
	SE	0.164	0.157	0.135	SE	0.329	0.32
	Control Mean	7.733	7.68	7.713	Control Mean	7.774	7.774
					T1 or T2 Effect	-0.229	0.001
					T3 Effect	0.09	0.09
	N	1998	1998	1998	N	1998	1998

RESULT 41: Feelings of safety in public spaces after dark among women

Description of measure: This measures the mobility of women within their home slum after dark (SAFETY_NIGHT). It is coded as the number of public places, from a list of 11 that women feel safe or very safe visiting alone after dark using MS4 A-K in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	-0.019	0.27	0.036	Interaction Effect	-0.262	-0.632*
	SE	0.189	0.182	0.153	SE	0.375	0.359
	Control Mean	3.579	3.465	3.48	Control Mean	3.341	3.341
					T1 or T2 Effect	0.112	0.586
					T3 Effect	0.336	0.336
	N	1996	1996	1996	N	1996	1996
Female Indirect	Treatment Effect	0.109	0.032	0.144	Interaction Effect	0.002	0.008
	SE	0.151	0.16	0.128	SE	0.286	0.302
	Control Mean	3.245	3.312	3.193	Control Mean	3.244	3.244
					T1 or T2 Effect	0.108	0.028
					T3 Effect	0.14	0.14
	N	1998	1998	1998	N	1998	1998

Qualitative findings on women's mobility and feelings of safety

In general, across the sample, women's mobility is severely restricted, they need permission from husbands to go out, and movement of women is only socially accepted – for limited purposes (i.e. linked to their role in the household). (See baseline for cross reference here.) Norms around women's mobility continue to impact on income earning opportunities.

Within this restrictive environment, improved mobility is seen as one of the most significant changes in the last 12 months by DBs in both T1 and T2 – which women link directly to their SHG membership. However, improvements in mobility seem to be limited to movement for SHG meetings – with women still often requiring permission to leave the home from their husband – even for SHG meetings, and limited activities for which it is socially acceptable for women to go outside of the home. Improvements in mobility are also limited to day time with mobility at night remaining restrictive for women (i.e. past 8/9pm at night).

> '*When I started going out, people started gossiping like "look at this woman, how she is roaming all around". They don't say like this anymore, my husband used to beat me for it, but now he doesn't do that'.* (Woman DB, T2 slum, Bhopal)

> *'Now it's like ma'am that if we want to go for group meeting our husbands don't deny, but if want to go anywhere else we cannot go without their permission. If we come late in night they say where were you at this time of night? If we say in group then they don't say anything.'* (Woman DB, T1, Jabalpur)

In line with the quantitative results, women are less afraid of travelling around their home slum during the day time, whereas at night time it is generally not considered safe for women to travel alone. However there do not appear to have been changes in women's perceptions of safety as a result of interventions.

Actions taken to address VAWG

This sub section includes results for all of the intermediate outcome measures relating to actions taken to address VAWG including:

- Support from SHGs to women who experience IPV
- Support from police to women who experience IPV
- Support from SHGs to women who experience public VAW
- Support from police to women who experience public VAW
- Actions taken to prevent or respond to IPV
- Actions taken to prevent or respond to public VAWG
- Recognition of the state as an actor to bring about social change
- Willingness to engage others around the issue of VAWG

Expectations of receiving support from SHG and police

Women were asked about their expectations of receiving support from their local SHG and/or police if they experienced IPV and/or public VAWG.

Overall women report expecting SHG members to be supportive if women experienced IPV or public VAWG and approached their SHG.

There is some evidence of treatment effects on expectations of SHG members for FDs in T2. In what is perhaps the strongest evidence for treatment effects in this study, FDs in T2 were about 10% more likely to report expecting support from SHGs. In terms of the range of the variable this corresponds to about one in five SHG members shifting from a "somewhat unlikely" position to a "somewhat likely" position. Furthermore, women are significantly more likely to expect their SHGs to be helpful following an incident of public violence; moreover there is some evidence for this effect operating also for indirect women beneficiaries, suggesting a moderate spill-over in confidence in SHGs to broader communities.(Result 42 and 43)

Women also report overall strikingly positive expectations of support from the police after experiencing IPV, and moderately positive expectations of police responding to reports of public violence, with most saying they expect support received would be somewhat helpful. However, there is no evidence of treatment effects on women's expectations of police. (Result 44 and 45)

Taking actions to prevent or respond to VAWG

Men and women were asked about how likely they would be to take a number of actions to prevent or respond to VAWG.

Both women and men report to be somewhat likely to take actions to prevent or respond to both IPV and public VAWG. There is no evidence of treatment effects on these outcome measures. The only exception is MIs in T3 slums who report a slightly lower likelihood of taking actions to prevent or respond to IPV. (Result 46 and 47)

Recognition of the state as an actor to bring about social change

As with claims around reporting violence to the state, respondents overall suggest that they recognise the state as a key actor for addressing IPV with most saying that the state is important or very important. There is variability across treatment arms but it is difficult to interpret: women suggest less of a role for the state in T3 areas; indirect women suggest less of a role for the state in T1 areas and men indirect beneficiaries suggest a greater role for the state in T3 areas. (Result 48)

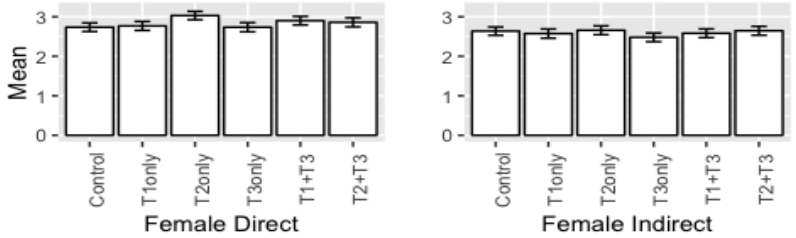
There is less variation in views about the importance of the state for addressing public VAW. MI again see a greater role in T3 areas but otherwise there is no clear effect. (Result 49)

Willingness to engage others around the issue of VAWG

Using data from the behavioural measure, about two thirds of women and a half of men report willingness to engage others in addressing violence against women by taking petition cards for others to sign. However, there is no evidence of treatment effects on women or men's willingness to engage others on the issue of VAWG using this measure.

RESULT 42: Receiving support from SHG/SHG member after experiencing IPV

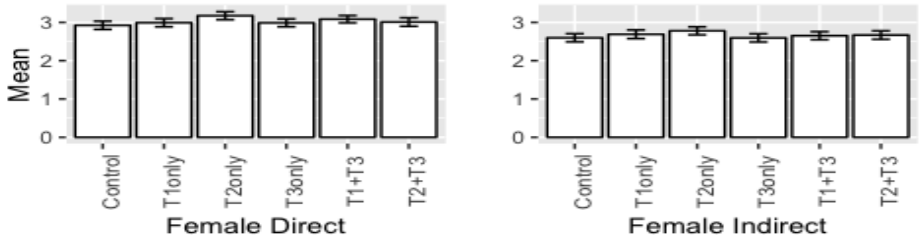
Description of measure: This measures the expected level of helpfulness of an SHG/SHG member upon reporting an incident of IPV, among married/cohabiting women (SIPV_SHG). It is coded as on a scale of 1-4, where 1 is very unhelpful and 4 is very helpful using RDV5 in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	0.083	0.198***	0.007	Interaction Effect	0.166	-0.172
	SE	0.053	0.056	0.044	SE	0.108	0.108
	Control Mean	2.841	2.787	2.844	Control Mean	2.736	2.736
					T1 or T2 Effect	-0.001	0.285***
					T3 Effect	0.009	0.009
	N	1753	1753	1753	N	1753	1753
Female Indirect	Treatment Effect	-0.01	0.037	-0.045	Interaction Effect	0.138	0.129
	SE	0.051	0.047	0.039	SE	0.104	0.098
	Control Mean	2.611	2.573	2.63	Control Mean	2.658	2.658
					T1 or T2 Effect	-0.08	-0.029
					T3 Effect	-0.135	-0.135
	N	1658	1658	1658	N	1658	1658

RESULT 43: Receiving support from SHG/SHG member after experiencing public VAW

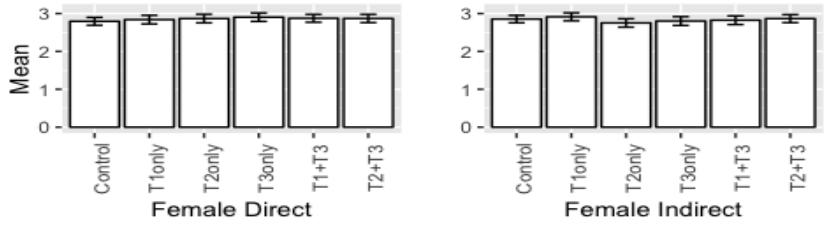
Description of measure: This measures the expected level of helpfulness of an SHG/SHG member upon reporting an incident of public harassment, among women (SPV_SHG). It is coded as on a scale of 1-4, where 1 is very unhelpful and 4 is very helpful using RPV4 in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	0.08	0.134**	0.005	Interaction Effect	0.081	-0.18
	SE	0.052	0.055	0.043	SE	0.109	0.112
	Control Mean	3.024	2.997	3.03	Control Mean	2.923	2.923
					T1 or T2 Effect	0.04	0.225**
					T3 Effect	0.038	0.038
	N	1994	1994	1994	N	1994	1994
Female Indirect	Treatment Effect	0.045	0.096*	-0.053	Interaction Effect	-0.021	-0.084
	SE	0.053	0.056	0.044	SE	0.109	0.113
	Control Mean	2.665	2.637	2.695	Control Mean	2.612	2.612
					T1 or T2 Effect	0.056	0.139*
					T3 Effect	-0.018	-0.018
	N	1988	1988	1988	N	1988	1988

RESULT 44: Receiving support from police after experiencing IPV

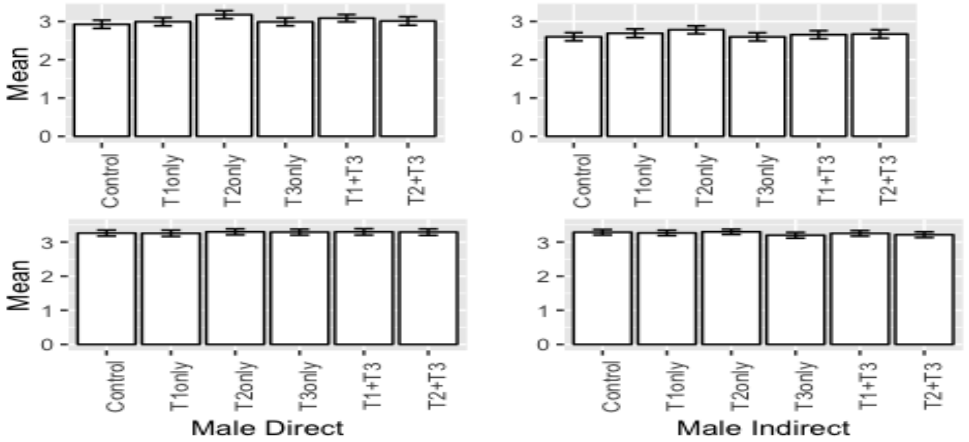
Description of measure: This measures the expected level of helpfulness of the police upon reporting an incident of IPV, among married/cohabiting women (SIPV_POLICE). It is coded as on a scale of 1-4, where 1 is very unhelpful and 4 is very helpful using RDV10 in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	-0.008	0.011	0.045	Interaction Effect	0.003	-0.076
	SE	0.048	0.046	0.04	SE	0.099	0.09
	Control Mean	2.858	2.853	2.834	Control Mean	2.794	2.794
					T1 or T2 Effect	-0.01	0.049
					T3 Effect	0.07	0.07
	N	1753	1753	1753	N	1753	1753
Female Indirect	Treatment Effect	0.021	-0.035	-0.008	Interaction Effect	-0.026	0.157
	SE	0.049	0.048	0.038	SE	0.096	0.096
	Control Mean	2.819	2.849	2.84	Control Mean	2.856	2.856
					T1 or T2 Effect	0.034	-0.115
					T3 Effect	-0.053	-0.053
	N	1660	1660	1660	N	1660	1660

RESULT 45: Receiving support from SHG/SHG member after experiencing public VAW

Description of measure: This measures the expected level of helpfulness of an SHG/SHG member upon reporting an incident of public harassment, among women (SPV_SHG). It is coded as on a scale of 1-4, where 1 is very unhelpful and 4 is very helpful using RPV4 in the endline survey instrument.

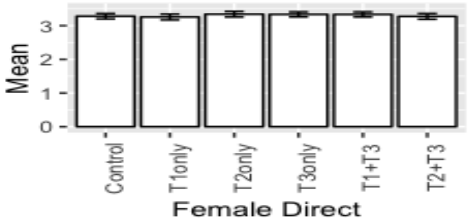


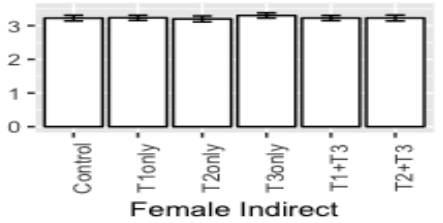
Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	0.08	0.134**	0.005	Interaction Effect	0.081	-0.18
	SE	0.052	0.055	0.043	SE	0.109	0.112
	Control Mean	3.024	2.997	3.03	Control Mean	2.923	2.923
					T1 or T2 Effect	0.04	0.225**
					T3 Effect	0.038	0.038
	N	1994	1994	1994	N	1994	1994
Female Indirect	Treatment Effect	0.045	0.096*	-0.053	Interaction Effect	-0.021	-0.084
	SE	0.053	0.056	0.044	SE	0.109	0.113
	Control Mean	2.665	2.637	2.695	Control Mean	2.612	2.612

					T1 or T2 Effect	0.056	0.139*
					T3 Effect	-0.018	-0.018
N	1988	1988	1988	N		1988	1988

RESULT 46: Taking actions to prevent or respond to IPV

Description of measure: This measures the average likelihood of taking 3 hypothetical actions to prevent or respond to IPV (ACT_IPV). It is coded as on a scale of 1-4, where 1 is very unlikely and 4 is very likely using IN3C-E in the endline survey instrument



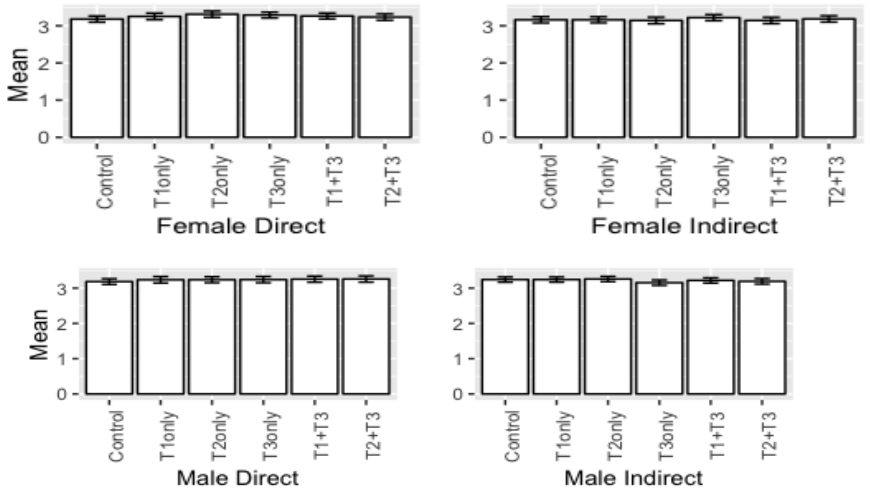


Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	-0.02	-0.012	0.021	Interaction Effect	0.092	-0.1
	SE	0.041	0.041	0.034	SE	0.082	0.082
	Control Mean	3.307	3.301	3.293	Control Mean	3.28	3.28
					T1 or T2 Effect	-0.066	0.038
					T3 Effect	0.024	0.024
	N	1995	1995	1995	N	1995	1995
Female Indirect	Treatment Effect	-0.024	-0.044	0.027	Interaction Effect	-0.078	-0.048
	SE	0.039	0.044	0.033	SE	0.078	0.087
	Control Mean	3.237	3.246	3.218	Control Mean	3.224	3.224
					T1 or T2 Effect	0.016	-0.02
					T3 Effect	0.069	0.069
	N	1998	1998	1998	N	1998	1998
Male Direct	Treatment Effect	-0.022	-0.011	0.048*	Interaction Effect	0.053	-0.047
	SE	0.03	0.03	0.025	SE	0.063	0.061

	Control Mean	3.286	3.277	3.273	Control Mean	3.263	3.263
					T1 or T2 Effect	-0.049	0.013
					T3 Effect	0.046*	0.046*
	N	1498	1498	1498	N	1498	1498
Male Indirect	Treatment Effect	0.005	-0.01	-0.052**	Interaction Effect	0.058	-0.024
	SE	0.03	0.03	0.024	SE	0.058	0.059
	Control Mean	3.25	3.252	3.285	Control Mean	3.286	3.286
					T1 or T2 Effect	-0.024	0.002
					T3 Effect	-0.063**	-0.063**
	N	1988	1988	1988	N	1988	1988

RESULT 47: Taking actions to prevent or respond to public VAW

Description of measure: This measures the average likelihood of taking 3 hypothetical actions to prevent or respond to public harassment or violence against women (ACT_PV). It is coded as on a scale of 1-4, where 1 is very unlikely and 4 is very likely using IN3A-B and IN3F in the endline survey instrument.

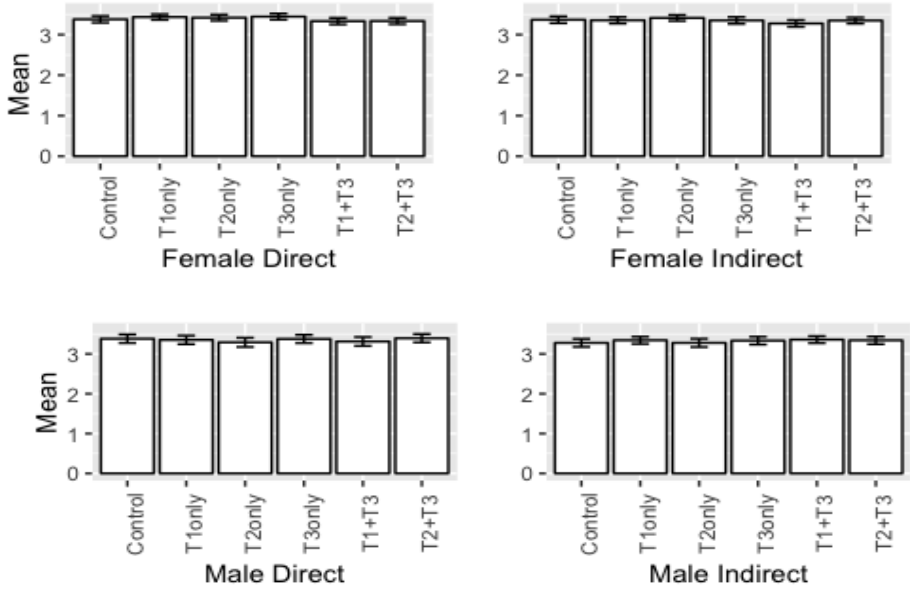


Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	0.006	0.025	0.013	Interaction Effect	0.013	-0.175**
	SE	0.04	0.039	0.034	SE	0.077	0.08
	Control Mean	3.252	3.245	3.247	Control Mean	3.179	3.179
					T1 or T2 Effect	0	0.112
					T3 Effect	0.068	0.068
	N	1996	1996	1996	N	1996	1996

Female Indirect	Treatment Effect	-0.041	-0.022	0.021	Interaction Effect	-0.036	-0.001
	SE	0.038	0.039	0.032	SE	0.078	0.08
	Control Mean	3.176	3.17	3.154	Control Mean	3.16	3.16
					T1 or T2 Effect	-0.023	-0.022
					T3 Effect	0.034	0.034
	N	1998	1998	1998	N	1998	1998
Male Direct	Treatment Effect	0.02	0.009	0.047*	Interaction Effect	0.026	-0.051
	SE	0.033	0.034	0.027	SE	0.067	0.065
	Control Mean	3.232	3.231	3.222	Control Mean	3.186	3.186
					T1 or T2 Effect	0.007	0.034
					T3 Effect	0.056*	0.056*
	N	1498	1498	1498	N	1498	1498
Male Indirect	Treatment Effect	0.028	0.017	-0.057**	Interaction Effect	0.096	0.02
	SE	0.029	0.027	0.024	SE	0.058	0.055
	Control Mean	3.213	3.216	3.249	Control Mean	3.244	3.244
					T1 or T2 Effect	-0.02	0.007
					T3 Effect	-0.096**	-0.096**
	N	1988	1988	1988	N	1988	1988

RESULT 48: Recognition of the state as an actor to bring about change in IPV

Description of measure: This measures the average importance of national, state and municipal governments for addressing IPV (IPV_STATE). It is coded as on a scale of 1-4, where 1 is completely irrelevant and 4 is very important using BH13A-C in the endline survey instrument.

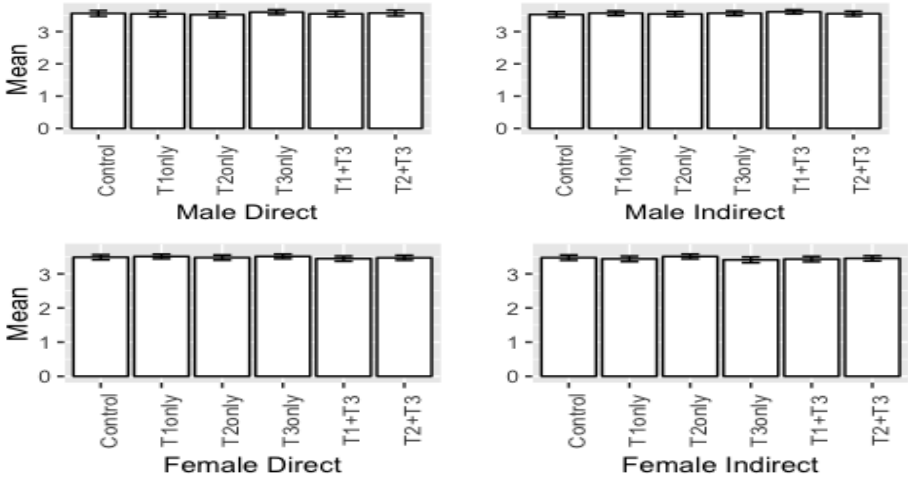


Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	-0.043	-0.031	-0.06**	Interaction Effect	-0.131**	-0.142**
	SE	0.032	0.031	0.026	SE	0.065	0.063
	Control Mean	3.394	3.397	3.411	Control Mean	3.378	3.378

					T1 or T2 Effect	0.022	0.04
					T3 Effect	0.031**	0.031**
	N	1996	1996	1996	N	1996	1996
Female Indirect	Treatment Effect	-0.078**	0.005	-0.056*	Interaction Effect	-0.014	0.002
	SE	0.035	0.036	0.029	SE	0.07	0.071
	Control Mean	3.369	3.337	3.378	Control Mean	3.367	3.367
					T1 or T2 Effect	-0.072**	0.005
					T3 Effect	-0.052*	-0.052*
	N	1998	1998	1998	N	1998	1998
Male Direct	Treatment Effect	-0.047	-0.078*	0.038	Interaction Effect	0.038	0.128
	SE	0.045	0.045	0.036	SE	0.088	0.088
	Control Mean	3.359	3.355	3.34	Control Mean	3.381	3.381
					T1 or T2 Effect	-0.066	-0.141*
					T3 Effect	-0.018	-0.018
	N	1498	1498	1498	N	1498	1498
Male Indirect	Treatment Effect	0.032	-0.038	0.058**	Interaction Effect	0.028	0.039
	SE	0.036	0.036	0.029	SE	0.071	0.072
	Control Mean	3.307	3.328	3.298	Control Mean	3.276	3.276
					T1 or T2 Effect	0.018	-0.057
					T3 Effect	0.036**	0.036**
	N	1987	1987	1987	N	1987	1987

RESULT 49: Recognition of the state as an actor to bring about change in public VAW

Description of measure: IPV_STATE measures the average importance of national, state and municipal governments for addressing public VAW. It is coded as on a scale of 1-4, where 1 is completely irrelevant and 4 is very important using BH14A-C in the endline survey instrument.

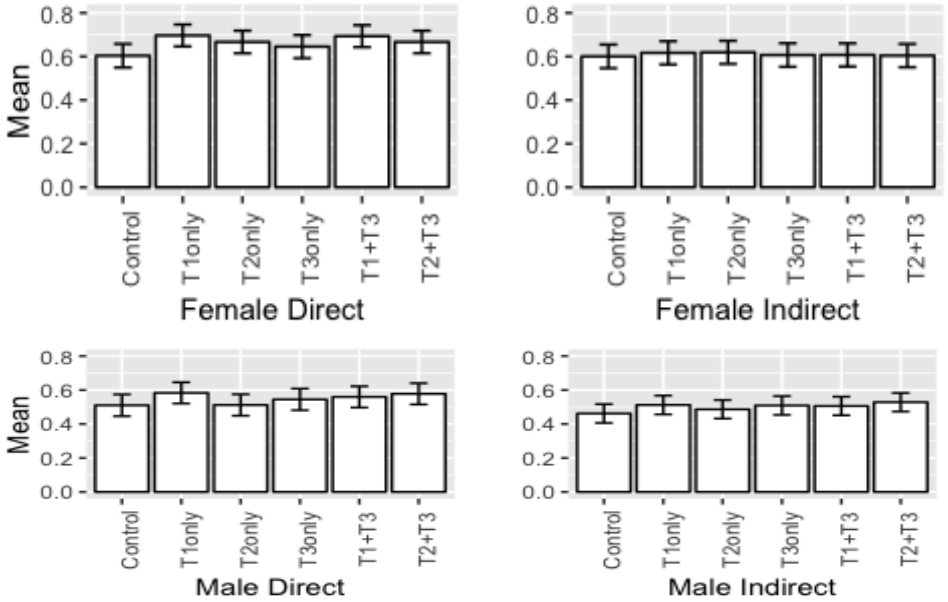


Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	-0.034	-0.022	-0.037	Interaction Effect	-0.028	-0.026
	SE	0.03	0.028	0.023	SE	0.065	0.057
	Control Mean	3.484	3.485	3.488	Control Mean	3.485	3.485
					T1 or T2 Effect	-0.02	-0.009
					T3 Effect	-0.018	-0.018
	N	1996	1996	1996	N	1996	1996
Female Indirect	Treatment Effect	-0.04	0.022	-0.054*	Interaction Effect	0.133*	0.029
	SE	0.034	0.03	0.028	SE	0.071	0.062
	Control Mean	3.459	3.436	3.472	Control Mean	3.472	3.472
					T1 or T2 Effect	-0.106	0.007

					T3 Effect	-0.108*	-0.108*
	N	1998	1998	1998	N	1998	1998
Male Direct	Treatment Effect	-0.022	-0.052	0.047	Interaction Effect	-0.021	0.036
	SE	0.041	0.039	0.03	SE	0.078	0.073
	Control Mean	3.565	3.565	3.546	Control Mean	3.563	3.563
					T1 or T2 Effect	-0.011	-0.069
					T3 Effect	0.042	0.042
	N	1498	1498	1498	N	1498	1498
Male Indirect	Treatment Effect	0.028	-0.026	0.043	Interaction Effect	-0.002	-0.011
	SE	0.033	0.034	0.027	SE	0.064	0.068
	Control Mean	3.549	3.569	3.548	Control Mean	3.529	3.529
					T1 or T2 Effect	0.029	-0.021
					T3 Effect	0.048	0.048
	N	1987	1987	1987	N	1987	1987

RESULT 50: Willingness to engage others in the community on VAW

Description of measure: This measures the willingness to engage others in the community on the issue of VAW (TAKE_ENV). It is coded as 1 if the respondent accepts a set of petition-cards to distribute among community members using BH12 in the endline survey instrument.



Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	0.062**	0.037	0.021	Interaction Effect	-0.028	-0.009
	SE	0.026	0.026	0.02	SE	0.049	0.05
	Control Mean	0.646	0.66	0.656	Control Mean	0.604	0.604
					T1 or T2 Effect	0.076**	0.042
					T3 Effect	0.033	0.033
	N	1996	1996	1996	N	1996	1996

Female Indirect	Treatment Effect	0.012	0.004	0.003	Interaction Effect	0.019	0.013
	SE	0.025	0.025	0.021	SE	0.051	0.051
	Control Mean	0.608	0.608	0.612	Control Mean	0.601	0.601
					T1 or T2 Effect	0.002	-0.003
					T3 Effect	-0.008	-0.008
	N	1998	1998	1998	N	1998	1998
Male Direct	Treatment Effect	0.064**	0.038	0.017	Interaction Effect	-0.04	0.018
	SE	0.029	0.027	0.023	SE	0.059	0.051
	Control Mean	0.536	0.55	0.535	Control Mean	0.51	0.51
					T1 or T2 Effect	0.084**	0.029
					T3 Effect	0.025	0.025
	N	1498	1498	1498	N	1498	1498
Male Indirect	Treatment Effect	0.026	0.027	0.023	Interaction Effect	-0.042	-0.028
	SE	0.024	0.028	0.022	SE	0.049	0.056
	Control Mean	0.497	0.497	0.487	Control Mean	0.462	0.462
					T1 or T2 Effect	0.046	0.041
					T3 Effect	0.046	0.046
	N	1987	1987	1987	N	1987	1987

7. Heterogeneous effects

Even if there is no evidence for overall treatment effects it is still possible that interventions were effective in some subgroups, or possibly had positive effects for some groups and negative effects for others. For example there could be differences across cities reflecting different quality of implementation or different background conditions. There could also be differences depending on the caste or religion of respondent, or past measured exposure to violence.

To assess heterogeneous effects across cities (or other categories) models and reporting that differ in four ways from the main analysis regressions and tables are applied:

1. The models include interactions between cities and treatments but they exclude treatment interaction effects (T1*T3 and T2*T3). This facilitates interpretation and calculation of fitted estimates in each city.
2. The average effect of each treatment for each subgroup with associated *p* values for these subgroup effects is reported.
3. In the case of heterogeneous effects other than city effects, a row labelled "Interaction effect" is included with an estimate of the interaction between the category of interest and the treatment (T1, T2, or T3). This is generally the difference between the average effect in the two conditions.
4. For all heterogeneous effects tables, the final column labelled "Pr(>F)" is the p-value of a statistical test to examine whether we can reject the null hypothesis that *all* of the interaction terms between treatments and the relevant heterogeneous variable are equivalent to zero (i.e. the models with and without the inclusion of these interaction terms are equivalent). If Pr(>F) is less than .05, the test is considered statistically significant at the 95% level and we can reject the null hypothesis that there is no heterogeneity. Low *p*-values in this column indicate places where heterogeneous effects may be important to note.

Figure 5 below shows the distribution across outcome measures of the *p* values from the hypothesis test that there are no heterogeneous effects across cities. If there were no heterogeneous effects across cities on any measure then the expectation is that this distribution would be flat. In fact it is left skewed. This suggests that there is evidence for heterogeneity in effects, though it does not imply that there are positive effects---or even significant effects---in any particular city.

Figure 5: Distribution across outcome measures – heterogeneous effects

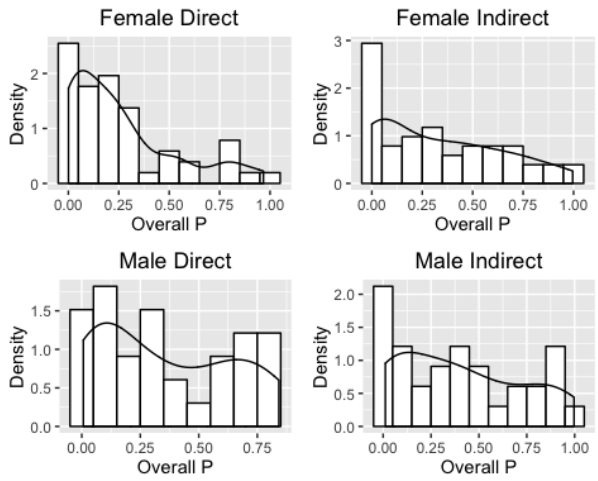
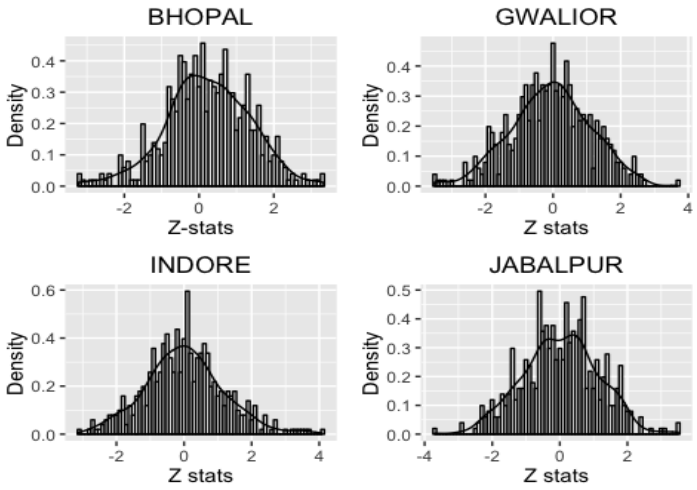


Figure 6: Heterogeneous effects across cities



Examining the patterns measure by measure suggests that on some items there is some stronger evidence for positive effects in Bhopal. In general, treatment effects in Gwalior and Jabalpur are somewhat more often out of line with programme goals. Figure 6 shows the distribution of *p* values across outcomes, treatments, and subgroups, by city. Though difficult to see, the share of significant values is marginally higher for Bhopal (9.5%) and Indore (9.7%) and lowest for Jabalpur (7.7%). These differences are not great however and in all cases the overall distribution of test statistics is close to normal.

While there is a lack of rigorous quantitative implementation data at the slum level that can be integrated methodically with the quantitative analysis, it is plausible that some of the differences in city effects are related to differences in implementation. Notably, the IP have suggested that implementation was more robust in Bhopal and slightly more robust in Indore than in Gwalior and Jabalpur. There has also been some qualitative evidence suggesting harmful programme effects in Gwalior and suggestions about issues with implementation. The IP have also discussed the wide variation even in programming such as strategies to involve male direct beneficiaries.

The heterogeneous effects results are largely consistent with these implementation factors. While we cannot say definitively that differences are due to implementation differences, we can say from this analysis that there is significant room for variation in the effects of a VAW programme of this type when applied across different locations.

Annex 7 presents all of the heterogeneous results data presented in tables by outcome measure.

PART C: CONCLUSIONS

This evaluation assessed the impact of the SCI on levels of VAW as well as intermediate outcome areas. The evaluation was designed to not simply generate evidence of any effects which can be attributed to the Programme overall, but more specifically to identify which of the Programme's interventions (SHG strengthening, SHG+VAW and Life skills education with men and boys) are most effective – and why.

Consistency of findings

Part B presented a significant number of results. The factorial design allowed us to assess a variety of programme types and combinations. The quantitative component of this evaluation finds little evidence that any of the treatments included in the SCI or combinations of them had positive impacts on the hypothesised outcomes.

Note that the quantitative analysis examined the effects of the Programme on 61 outcome variables for between two and four different populations (including both direct and indirect, male and female beneficiaries) drawn from each slum, resulting in a large number of statistical tests. Thus, when analysing the effectiveness of the various treatment arms, it is important to focus on the consistency of findings across outcomes rather than on any particular result.

Overall, there are a very small number of Programme effects on particular outcomes, and over half of the effects are not in the hypothesised direction. Furthermore, results are most limited where we would most expect to see them – at the level of the direct beneficiaries. Given the focus on the interventions to reduce social tolerance of VAWG, the almost consistent lack of Programme effects for direct beneficiaries on intermediate outcomes such as individual attitudes (with the exception of the men and boys intervention on attitudes towards IPV) is particularly striking.

Findings from the qualitative analysis on outcomes of interest highlight wider perceived benefits of SHG membership including improved social networks and confidence as well as some signs that key messages in the Life Skills module may have been effective in encouraging men and boys to challenge unequal gender roles in the household. However, there is limited evidence in wider shifts in attitudes, norms and behaviours from the qualitative data, which is broadly supportive of the quantitative findings. In particular, the qualitative data highlights persistent harmful norms and attitudes which continue to drive and sustain VAWG in urban slums in Madhya Pradesh.

Taking into account the consistency of results across the outcome areas we conclude that there is little robust evidence that any of the interventions – SHG strengthening (T1), SHG+VAW (T2) or Life Skills curricula (T3) with men and boys – led to improved outcomes.

Plausible explanations for the results

The factorial design allowed us to assess a variety of programme types and combinations and our analysis suggests that **the null results are not associated with the any of the three interventions**, however the statistical evidence is not able to tell us why the Programme failed to achieve its' intended effects.

In order to account for the evaluation findings, and to generate learning and insights to inform future VAWG programmes and evaluations, three possible explanations for the null results are explored:

1. **Implementation failure -** i.e. Due to challenges in implementation, the SCI was not implemented as intended, there was a lack of Programme fidelity. This explanation suggests that underlying issues were not associated with application specific to this context, but simply an issue of delivery.

2. **Theory failure** – i.e. The theory of change associated with this programme does not reflect how change happens on the ground. This means that the Programme would not have worked even if implemented perfectly.
3. **Measurement error** - i.e. The research failed to reveal the true results, because the design, the measurement or the analysis was weak.

While the quantitative component of this study was not designed to differentiate between these three possibilities, we provide information based on monitoring data, qualitative data, the accounts provided by implementers, and wider experience and evidence in the field of VAWG prevention.

While a central goal of this evaluation was to understand the effectiveness of specific VAWG prevention programmes in bringing about change, questions about measurement of sensitive measures and implementation make it difficult to pinpoint whether null results are due to the ToC, implementation, or even measurement.

The following sub-sections explore each of these three plausible explanations for the null results in more detail drawing from the quantitative and qualitative data, monitoring data and correspondence with the IP.

8. Diminished implementation fidelity

Assessing implementation fidelity – the degree to which an intervention is delivered competently and according to protocol – is critical to interpreting the results of the evaluation. The available evidence on delivery suggests a number of areas which are likely to have adversely affected the ability of the Programme to achieve its' intended results.

Programme duration

The SCI was designed to be delivered over a two-year period from January 2014 to December 2015. However, due to significant delays in the procurement of CSAs, the agencies were not formally appointed until up to nine months after the planned start date. Furthermore, field activities were constrained due to the imposition of three Election Codes of Conduct in 2013 and 2014 which interrupted delivery for 2 months. Consequently, the implementation period across the four cities varied from approximately 9 to 15 months – with the longest delivery time in Jabalpur and Bhopal and the shortest in Gwalior and Indore.

The overall project duration is longer than the intervention delivery period in each slum, as core activities were effectively staggered by slum and intervention type at the city level. As a result, in each city core intervention activities were not delivered in some slums until the last 3-6 months of the implementation phase. [61]

It seems reasonable to conclude that the Programme may simply have been too short to expect measurable changes to occur in deep seated attitudes, norms and behaviour. The 9-15-month duration and even the two years intended duration - may not have been sufficient to achieve impact level changes in social norms and prevalence of VAWG. However, one might still have expected some positive change in intermediate outcomes such as individual attitudes within shorter timeframes.[62]

Opportunities for piloting and adaptive programming

A longer timeframe may have provided an opportunity to pilot the approach ahead of implementation, to allow for a longer community mobilisation period, to have ensured CSA capacity was sufficient and to have delivered a longer training to peers from SHGs and men and boys groups – to improve their capacity to

[61] This was also in part because of delays in payments– with some CSAs reportedly not receiving the funding they needed to start delivery until June 2015.

[62] Fulu et al (2014)

facilitate monthly sessions (in line with the programme design – see page 111). These factors would likely have led to a more effective approach.

The short duration also meant that the sequencing of activities of the SCI was dictated by Programme delivery rather than the group or community's readiness to change.[63] This meant that in practice there was no opportunity to revisit a monthly theme, if the messages had not been well understood or attendance was particularly low.

The final SCI design was not piloted prior to implementation, so many of the mobilisation and logistical challenges encountered were not foreseen. Whilst adaptation to local contexts is necessary in VAWG prevention programming, the Programme lacked the flexibility to make significant well-considered and evidence-based adaptations in response to learning and changing circumstances due to the relatively short timeframe and requirements of the RCT.

Dosage of interventions

Another possible explanation for the disappointing results is that the 'dosage' of the interventions– the amount of intervention activities that beneficiaries were exposed to - may not have been sufficient to achieve changes in attitudes, norms and behaviours. Although there is no well-established minimum dosage for VAWG prevention interventions,[64] emerging learning from the What Works Programme indicates that ensuring there are enough intervention hours is likely more important than the length of project.[65]

As outlined in the implementation section, all three Modules followed a similar pattern of structured group training, followed by regular facilitated group meetings focussing on key messages. The frequency of group meetings varied by intervention type, with the SHG only slums (T1) receiving on average 10 monthly SHG sessions, men and boys (T3) receiving on average 11 sessions, and in SHG+VAW intervention slums (T2), SHGs receiving on average six monthly facilitated sessions. Sessions are assumed to have lasted on average one hour. Although the evidence base on optimal dosage is at an early stage of development, when comparing the SCI with other VAWG prevention programmes which have demonstrated shifts in attitudes, norms and behaviours, the SCI dosage appears to be lower than these other evaluated models.[66]

Furthermore, looking at dosage only in terms frequency of activities assumes high attendance, when evidence from the qualitative component as well as anecdotal data from the IP highlight significant difficulty in mobilising communities and low attendance at both structured trainings and regular meetings.[67] Unfortunately, the evaluation team were not given access to full monitoring records of slum level attendances, and so we are unable to verify this or calculate the average attendance at training events or facilitated sessions (see evaluation limitations). However, data on implementation checks (Part B section 5) on low recognition rates of the programme and core activities is supportive of low dosage. The available

[63] For example, a key quality of the approach of SASA! in Uganda is that ideas are introduced over time – using a stages of change model, based on the readiness of individuals and the community. Raising Voices, LSHTM and CEDOVIP. (2015).

[64] A key research question currently being explored by the DFID What Works Programme is what intensity/dosage is needed for effective programming.

[65] From correspondence with the South Africa Medical Research Council who lead a consortium including SDDirect to deliver the DFID What Works to Prevent VAWG programme.

[66] For example, the Stepping Stones and Creating Futures Model delivered in South Africa[66] - , requires 21 three-hour sessions. The SASA! approach in Uganda involves intensive regular mentoring with communities over a four-year period. Yaari Dosti adaptation of Program H in India delivered by the Population Council included weekly group discussions over six months.

[67] In FGDs, participants reported that some members did not attend the training events, and many participants reported to have not been able to attend meetings on a regular basis due to work commitments, other priorities and a lack of interest. Some FGD DB participants reported never having attended any of the group trainings or meetings.

evidence is worrying and supports the assertion that low dosage exacerbated by inconsistent attendance is likely to have been a significant contributory factor in explanation of the Programme results.

Capacity of partners

At the city level the Programme was implemented by ULBs with technical support from GHK/IPE's City Cluster teams (CCTs) and through CSAs in each city. Programme interventions were delivered and activities conducted through CSAs. In the ULBs, designated Nodal Officers and Community Development Officers monitored day-to-day activities with support from the CCTs.

CSAs were recruited by the ULBs through a tendering process that took account of quality and cost of services to be provided. Each CSA was to have a full time team leader, gender specialist and livelihood experts, part-time legal and microfinance experts and a counsellor, as well as 12 field workers to attend to field level activities in 52 slums. Unfortunately, none of the CSAs had a complete team of experts during Programme implementation. Furthermore, training and capacity building support was limited.

Content and design of curricula

Two curricula available in English were shared with the evaluation team. These comprised two linked (parts 1 and 2) for the SHG+VAW intervention (T2), as well as related key messages to be delivered through thematic sessions. The training was designed for the two animators selected from each SHG receiving the additional VAW intervention. As will be discussed below, this curriculum was then adapted for delivery to the members of all SHG+VAW groups, with key messages at regular thematic meetings aligning with the core content.

Whilst the animator level curriculum [68] reviewed by the evaluation team is well-conceived and well-structured, including good participatory exercises and clear instructions for facilitators, in parts it is didactic and there are several elements missing which would enable participants to develop practical skills to prevent and address violence in their families and communities:

- The SHG+VAW module (T2) curricula do not include practical sessions exploring the ways violence is triggered in intimate partnerships of families (e.g. infidelity or suspicions of infidelity, man not contributing economically, other economic pressures, refusal of sex etc.). Nor does it cover skills to improve communication around these topics and find non-violent alternatives to resolve issues (e.g. time out, compromising, talking, give and take). Equally, in the community, there is no exploration of practical skills for a woman to keep herself safe (e.g. routes to take, walking in pairs etc.).
- The SHG+VAW module (T2) curricula fail to tackle social norms as well as practical and economic reasons that might stop a woman from wanting or being able to go through each of the necessary stages to report violence to the police, apply for a protection order etc.
- The length of some sessions is likely to be too short and rushed for participants to absorb the information. There is no time built in for questions and answers and, aside from a short recap for 15 minutes at the start of the second day, there do not appear to be opportunities to revisit the material if messages are not well understood.

The thematic sessions were intended to impart key messages at regular intervals in line with the core curricula. The thematic sessions were structured around key messages, with flexibility for how these sessions were delivered pedagogically (see section below).

Delivery of curricula and trainings

According to the IP, these curricula were intended to be used as follows:

- Core team (IP) to train CSA staff on the curriculum and how to be an effective facilitator
- CSA staff to use the two curricula to train two animators per SHG on VAW
- The two animators were then to use / adapt these materials to communicate to the wider SHG and plan practical actions (community events, safety audits, meetings with authorities etc.).

The content of the animator training is complex and the interactive exercises need experienced facilitation. However, as a proposed methodology, there are already a number of challenges. This kind of train the trainers approach is often much less effective than foreseen, with significant losses in quality along the chain. This is even more so when using written materials, yet many of the end beneficiaries – and even the frontline trainers - are likely to have been illiterate or semi-literate. Important nuances and messages can be lost, and at worse, key parts if the curriculum are misinterpreted or misunderstood. As a result, the transmitted messages can produce harmful consequences.

Furthermore, from project documents, available monitoring reports and conversations with the IP, it seems that in practice the trainings were not delivered as planned due to time pressures and that further losses of quality are likely to have occurred:

(i) Training CSA staff: It seems that there was only a very limited training planned for the CSA staff on the three core curricula. This comprised a one-day orientation on the overall programme design, two or three days' training on the financial literacy module (SHG only), one day on the men and boys youth ambassador training (Men and Boys) and three days on the Gender+VAWG animator training (SHG+VAW). According to the IP, the animator curriculum was explained and some materials demonstrated, but this is unlikely to have been adequate given the complexity of the content. Thus, already at this stage it is likely that some of the quality was lost and it is possible that CSA staff who then delivered the training were themselves not progressive in their gender attitudes and behaviours.

(ii) Training animators: According to the IP, the animator training was implemented as per the curriculum by CSA staff. However, subsequently it was felt that the capacity of the animators to deliver the thematic sessions was not sufficient. A review of the SHG+VAW animator level training curriculum reveals that, although the material is generally well-conceived, it is complex to understand and the exercises are good but require a knowledgeable and experienced facilitator. The capacity of field workers to deliver trainings was a key challenge raised by the IP. This fact, coupled with the limited training and capacity building provided to CSA staff, means that it is highly likely that that CSA staff delivered this curriculum in a more didactic manner that prioritised teaching the key messages rather than promoting interactive discussion and reflection.

(iii) Training SHG members/men and boys: After concluding the animator trainings, the IP concluded that the animators would not have adequate capacity to then deliver further training and key messages to the rest of their group. A decision was therefore taken that, instead of delivering further training, the CSA staff would undertake training all SHG members/youth groups themselves (a key departure from the original programme design). Again it seems, because of time and budget constraints, they were asked to deliver a condensed version of the SHG+VAW training with only three days of training for the VAWG module (rather than five days) and women expected to take actions against VAWG in the following two days. For the Men and Boys Module, the youth group gender training was to be delivered over just one day (rather than three days). There are a number of potential ways in which this could have adversely affected quality:

a. Firstly, the training was not delivered by SHG members/men and boys' peers, but by outsiders (CSA staff) which might have affected the receptiveness of the women to the messages.

b. Secondly, the curriculum was condensed and delivered in three days rather than five days for the SHG+VAW and one day rather than three days for the men and boys training. This would have affected both the amount and quality of content and the capacity of participants to absorb key messages. Again it is likely that many of the participatory exercises were skipped.

c. Thirdly, the IP reports that the decision on how to condense and adapt the curriculum to the shorter time window and the full SHG group was left to CSA partners and there were no standardised adapted curricula available, nor training provided to CSAs on how to deliver the adapted curricula to target groups. As these group trainings took place at the slum level the IP was unable to assure consistent quality of delivery (see below). Thus, the evaluation team is unable to confirm which parts were taught and not taught, whether material was delivered in a rush or without using the practical exercises and opportunities for recap and how (in)consistent this was across slums or cities. Although there is a calendar record of what training was done on what dates (in the Project Closure Report), it seems there are no records of how the curriculum was adapted and delivered by different CSA staff members in different slums and in different cities.

Following delivery of the initial training, staff from the CSA in each city held follow-up thematic sessions with the SHGs and men's and boys' groups in each slum. These sessions were seen as an important way of ensuring key messages were absorbed and participants were able to reflect on their experiences and behaviours. There were intended to be 10 monthly sessions in total. In practice, in the slums with the SHG+VAW intervention, SHGs received an average of only six monthly facilitated sessions – thus, the intensity of these follow up sessions was lower than planned. The exception is the men's and boys' groups who on average had 11 monthly sessions. However, these were large group meetings and events rather than the smaller group work initially planned.

Furthermore, in order to adapt to the differing capacity of CSAs and contexts in the slums, the IP allowed the CSAs a certain amount of flexibility as to how they conveyed key messages after the training. There was no curriculum for this; CSAs were provided with the messages plus reading materials and films to use to animate sessions. This effectively meant that they were not required to organise structured sessions, but instead could use other means, such as large-scale slum events and leaflet distribution, to get across key messages on a regular basis.[69] This shift represents a fundamental diversion from the Programme design, and whilst evidence of small group learning is promising[70], evidence suggests large-scale awareness raising events may compound harmful social norms by reinforcing negative norms and making harmful behaviours more visible.[71]

This kind of curriculum needs very knowledgeable, skilled and experienced facilitators to deliver it well. It seems likely that CSA staff did not possess this background and, moreover, it seems that their training was inadequate. The Programme documentation and evidence from the qualitative component suggest that the pedagogic approach was weak with the emphasis on providing information and raising awareness rather than facilitating a process of critical reflection and learning. Coupled with the poor attendance at training sessions and a reduced number of follow-on mentoring sessions with no specific structure, this points to very weak

[69] In particular, it was deemed to be much easier to mobilise men and boys through large-scale events rather than small facilitated monthly discussions.

[70] See Fulu et al (2014)

[71] Alexander-Scott et al (2015) Shifting social norms to tackle violence against women and girls (VAWG). DFID Guidance notes

implementation of the curriculum. This 'implementation weakness' is likely to be the primary reason for the weak outcomes of these trainings in terms of change in attitudes and behaviours.

Mobilising communities

Implicit in the Programme logic and choice of target groups is the assumption that a) it would be possible to mobilise SHGs and men's and boys' groups around the issue of VAWG and b) that these groups will have wider influence on their community.

In practice, the IP faced significant difficulty mobilising groups. Men and boys were particularly challenging to mobilise and adaptations had to be made to the Programme to include sports activities as an incentive. Although SHGs were pre-existing groups on paper, in practice, they varied significantly in terms of group cohesion and activeness and the IP faced challenges keeping these groups together. Challenges to mobilising men and boys appear to have been compounded by local contextual counter forces. In the FGDs, many MDs also reported being part of the Bajrang Dals, Rashtriya Swayamsevak Sangh and other right-wing Hindu paramilitary or vigilante groups. These groups were perceived by men and boys to be significant actors working towards a better society through 'moral policing' (including reports of beating up young unmarried couples for example). Unlike the men and boys' groups– these groups were perceived as having influence and respect from the community and authorities. Furthermore, whilst these counter forces likely posed context-specific challenges for implementation, engaging with members of such groups also raises real risks of backlash – which does not appear to have been factored into the Programme design.

9. Theory weakness

The results may also indicate possible weaknesses related to the theoretical basis underpinning the Programme as well as theories embedded within the Programme activities, in part due to the limited evidence available on what works to prevent violence against women and girls, especially at the time of designing the programme.

The SHG model of economic empowerment

The Programme theory assumes that SHGs are an effective means of empowering women. The ToC assumes that the SHG strengthening will lead to increased access to government and other livelihood schemes, which in turn would lead to an improvement in income and control over income. However, the results show no evidence that SHG membership led to improvements in livelihoods, income or control over income. In fact, the results show a statistically negative impact for SHG members on control over income for the SHG strengthening module. This finding is alarming and may reflect a tightening of control by husbands in response to women's SHG membership, and warrants further attention.

These findings bring into question the SHG model of empowerment, and add support Jakimow and Kilby's (2006) assessment that in practice the SHG model is focused on solving market failures, by emphasising credit and saving, rather than empowering women.

Local VAWG response

Effective response mechanisms - including provision of access to justice, legal protection and services for survivors – is an important element in communicating non-acceptability of VAWG and creating an enabling environment for VAWG prevention efforts.[72] The results show that low trust in police response and perceived (and in some cases experienced) inadequacies in local response mechanisms continue to create a significant barrier to reporting IPV and pubic VAWG, whilst supporting norms around perceived impunity for perpetrators.

[72] Jewkes (2014)

Whilst the SCI did engage local actors – for example police - in activities at the community level, there were no specific activities aimed at building the capacity of local response, and/or improving services for survivors. Consequently, the effectiveness of the SCI was undermined by continued low trust in the formal response system.

The weak formal response is reflected in the monitoring data from the IP on reporting and referrals of VAW, where the majority of 129 cases reported to CSA field staff received counselling and informal mediation services (58), and only a minority (27%) receiving police action or referred to protection officers. From the qualitative data, there are reports of police emphasising family reconciliation and mediation over safety of survivor and justice. Furthermore, we know little about the *quality* of counselling and mediation services provided by local counsellors, nor the longer term outcome of these services on levels of VAW.

In addition, the SHG+VAW intervention encouraged SHG members to provide informal mediation and counselling to individuals experiencing IPV. However, it is questionable whether the short-term training provided during the VAWG intervention sufficiently equipped SHG members with the skills necessary to counsel perpetrators and survivors. Whilst local mediation can be effective especially in the absence of formal response mechanisms, as a recent ICRW report explains, mediation often assumes joint responsibility for violence; assumes that both partners have equal power; may put the survivor and their family in danger; and assumes that both partners want to live without violence.[73]

Social norms approach

At the time of the design of the SCI, social norms theory in the VAWG field was relatively new and the IP did not use social norms theory to guide programme design. Since then, however, there have been significant advances in both applied social norms theory and field research drawing important insights for programming.[74]

Our findings support evidence from the field that when harmful social norms are operative, they can provide a 'brake on social change'[75]. When discriminatory norms are prevalent, raising awareness of rights and laws (key approaches of the SHG+VAW (T2) and Life Skills (T3) modules) is not sufficient to shift social motivations of human behaviour. Furthermore, our findings indicate that descriptive norms – i.e. what people think other people do, may be a bigger driver of VAWG than prescriptive norms – i.e. what people think others expect them to do. Therefore, in this context awareness raising about the prevalence of VAWG may exacerbate the very social norms the intervention intends to shift.[76]

From the qualitative component of the evaluation, the social norm that IPV is a family matter was highly pervasive across both control and treatment slums. This norm serves to sustain violence and acceptance of IPV, whilst clearly presenting significant barriers to VAWG response and on direct beneficiairies ability to take actions to address VAWG (a key output of the ToC). Furthermore, social sanctions – including shame, loss of reputation and honour, impact on mobility, impact on marriage prospects, and increased risk of violence – present significant barriers to reporting VAWG. These specific social norms underpinning VAWG in the context of urban slums in Madhya Pradesh were not diagnosed or specifically targeted through the intervention.

Theory of diffusion

[73] See Heilman et al (2016)

[74] See for example recent DFID guidance on addressing social norms in VAWG programming. Alexander-Scott et al (2016)

[75] Ibid. (2016) citing Heise (2011)

[76] Ibid. (2016)

The Programme theory for how the interventions would diffuse to wider members of slums (indirect beneficiaries) was untested, and in practice, and a number of factors hindered effective diffusion:

- Men and boys faced significant challenges especially related to actions on IPV;
- Husbands of SHG members were not engaged which limited Programme impact;
- Other than at large scale awareness raising events, there is little evidence that direct beneficiaries were adequately supported to share key messages with the wider community.

10. Measurement weaknesses

Measuring VAWG involves particular challenges stemming from the sensitivity of the topic. Even under ideal conditions respondents may not be willing to share information about experiences of violence with a stranger implementing a survey. In field conditions concerns around confidentiality, the intentions of enumerators, or about being overheard by others can loom large.

To help understand the reliability of the survey data, the evaluation team undertook a set of tests to examine the consistency across baseline and endline measures on fixed and changing characteristics.

Table 20 below shows the correlation between individual responses across the two survey waves, for a core set of measures, as well as the correlation between slum level means at baseline and endline for the same measures.

Table 20: Correlation between individual responses baseline-endline

Measure	Ind. Correlation	Slum Level Correlation
Age	0.931	0.964
Religion (=1 if Muslim)	0.865	0.983
Marital Status	0.730	0.617
Dwelling status (kutcha, pucca or semi pucca)	0.384	0.834
Alcohol Consumption among men (=1 if ever consumed)	0.187	0.302
Childhood exposure to violence (=1 if ever witnessed/experienced violence)	0.061	0.092
Perpetration of VAW among men (=1 if ever perpetrated)	0.078	0.133
Experience of IPV (=1 if ever experienced)	0.094	0.029
Experience of Public Harassment (=1 if ever experienced)	0.007	-0.009

While both individual and slum-level correlations are relatively high on fixed and demographic characteristics (though in the case of the dwelling status, still low) they are sometimes extremely low for sensitive behaviours. In some cases slum level correlations are higher than individual level correlations. This could arise for example if a sensitive behaviour (such as alcohol consumption) is prevalent in a slum and in some periods some individuals admit to it while in other periods other individuals do.

The low correlation in the IPV measure[77] over time is a cause for concern: poorly measured variables make it more difficult to estimate treatment effects (while "classical" measurement error on outcomes makes for noisier estimation it does not induce attenuation bias). Unfortunately, it is not possible to conclude from

[77] Note we used standard outcome measures in line with the DFID What Works Programme and very similar to standard DHS questions. See annex 3 for the endline survey instrument.

these correlations alone whether the weak correspondence between measures is because the measures are unreliable in baseline, endline, or both. It is possible, however, to make some progress by comparing the direct endline VAW measures with the indirect measures gathered using the list experiment.

As described above, because of prior concerns with weaknesses in standard measures for VAW, the survey included indirect measures using the list experiment approach described above. Steps were also taken to try to validate the list experiment measure.

To validate the list experiment measure, the approach used regressed the number of items reported in the list experiment on an indicator of whether the respondent received a list with a sensitive item and whether the respondent reported experience or perpetration of the same sensitive behaviour in the direct survey question plus the interaction of these two. One would expect direct reporting to positively predict reporting in the list experiment i.e. those individuals who directly report experiencing or perpetrating a behaviour should also report it in the list experiment and this should be seen by looking at the interaction between the long list indicator and the direct measure.[78]

There is a positive interaction in all list experiment measures and it is generally strongly significant (the exception is the case of the married men where the number of observations is low). See Table 21. The fact that the LongList measure is still large and significant in this validation check means that there are many respondents who report positively to the indirect measure that did not report positively to the direct measure. However the fact that the coefficients on LongList and LongList*Direct generally sum to less than 1 suggests that some individuals that responded positively to the direct measure are not responding positively to the indirect measure.

Overall these results are reassuring - while there is certainly error in measurement, the correspondence here suggests that the endline direct measure is informative and that the indirect measure picks up this same information in addition to information that is not captured by the direct measures.

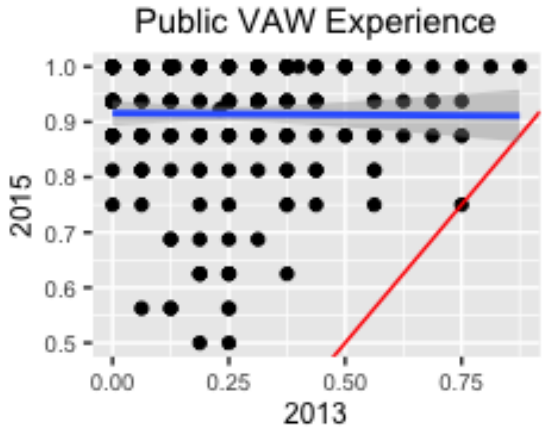
Table 21: Results from linear regression of list experiment on direct measure of same behaviour

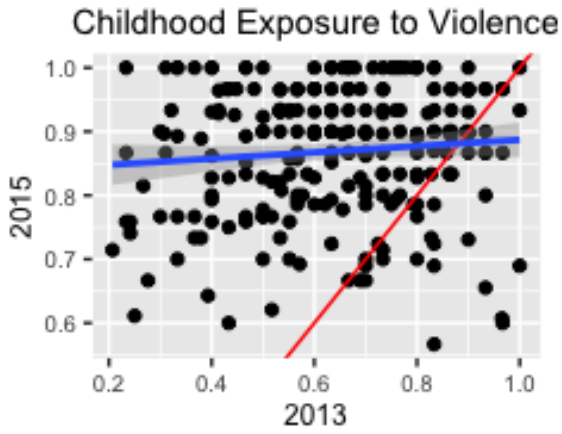
	IPV Experience	IPV Perpetration	PVAW Experience	PVAW Perpetration
Intercept	1.202***	1.311***	0.871***	0.959***
LongList[79]	0.187***	0.216***	0.104***	0.166***
Direct Report	-0.023	0.245	0.294***	0.307***
LongList*Direct	0.219*	0.437	0.3***	0.465***
Sample	Married Women	Married Men	All Women	All Men

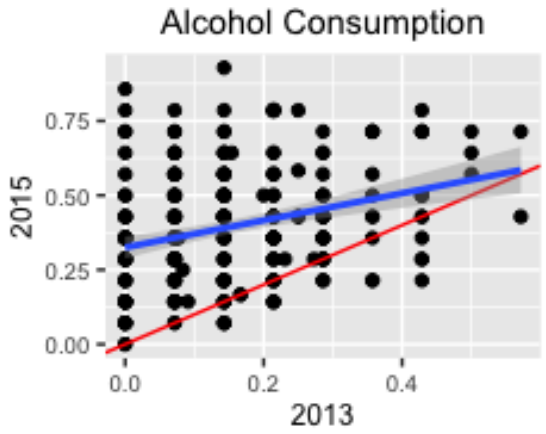
[78] More formally we estimate a model of the form "Number of items = *a* + *b** LongList + *c**Direct Report + *d** LongList *Direct Report", where LongList is a variable for whether the list included the sensitive item (VAW). In this model if all respondents answering positively in the direct measure also respond positively to the long list then *b*+*d* should take a value of 1. Coefficient *b* then captures the level of VAW for those that do not report positively to the direct question. Say then that 20% of subjects experience VAW, say half of these report positively in the direct question (and none that did not experience VAW report VAW in the direct question), then we would expect, in the absence of measurement error: Number of items = a + (1/9)* LongList + 0*Direct Report +(8/9)* LongList *Direct Report, for a total outcome of (9/10)* (1/9) + (1/10)*((1/9) + (8/9)) = (2/10).

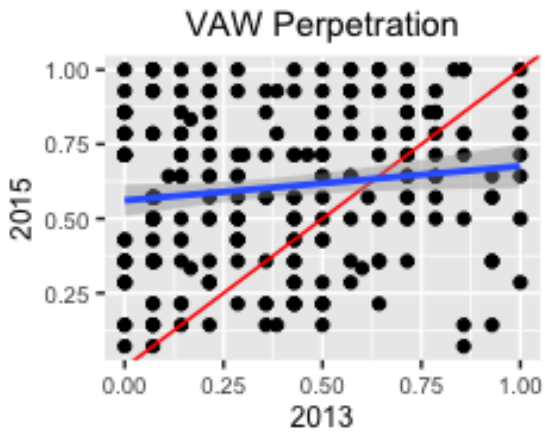
[79] For the list experiment we report the estimate of the average outcome in the control group, which is given by the coefficient on a variable that indicates whether a respondent received the long list (labelled "LongList" or "LongList Effect" in the tables").

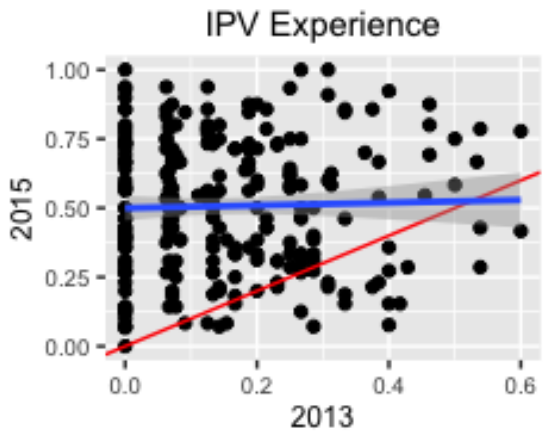
Observations	3418	1327	3993	3011

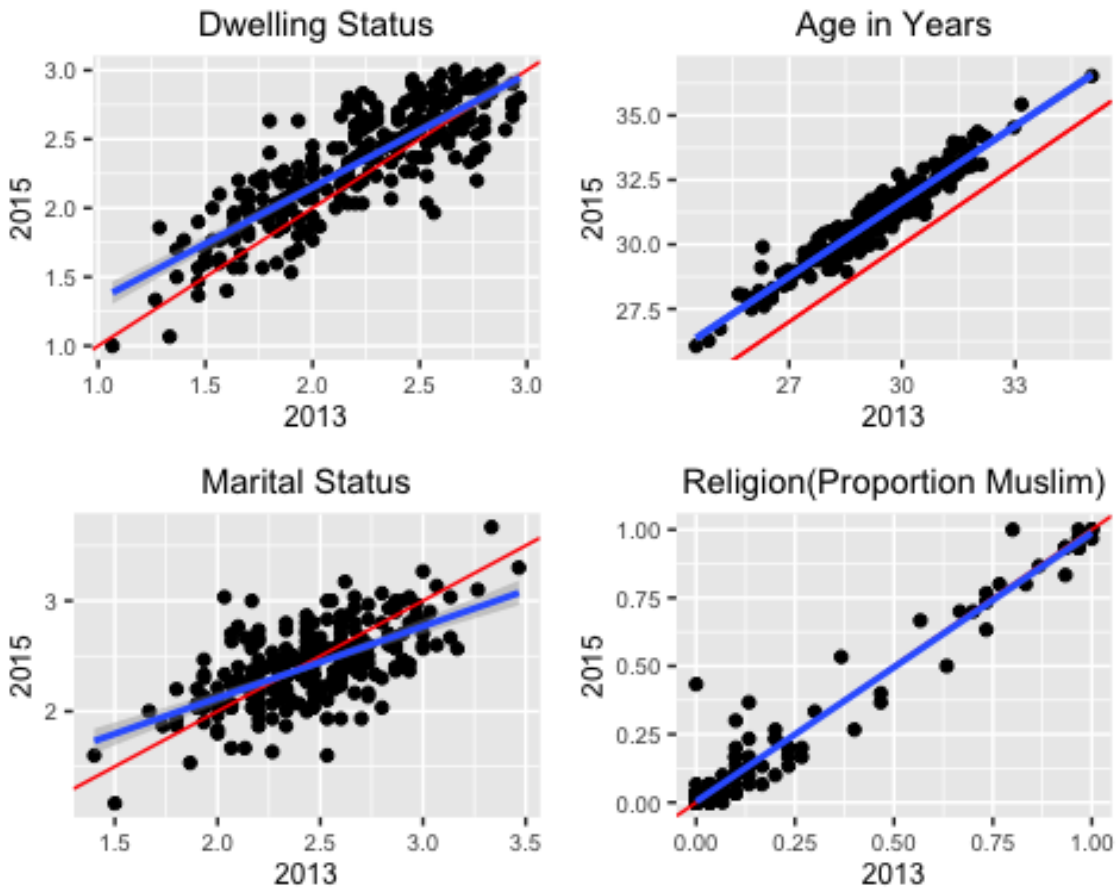











Dwelling Status
2015
2013
3.0
2.5
2.0
1.5
1.0
Age in Years
35.0
32.5
30.0
27.5
27
30
33
Marital Status
3
2
3.5
Religion(Proportion Muslim)
1.00
0.75
0.50
0.25
0.00

PART D: INSIGHTS FOR FUTURE VAWG PROGRAMMES AND EVALUATIONS

While the statistical analysis provides no clear explanation for the results, by comparing what we know about the SCI with evidence and experience in the wider VAWG prevention field, the SDDirect team offers the following practical insights for future VAWG prevention programming and evaluations.

10. Insights for VAWG programmes

Achieving transformational change at scale requires higher intensity and consistent programming

The SCI represents an attempt by DFID and the Government of Madhya Pradesh to integrate a VAWG component onto an existing large-scale urban infrastructure programme (the MPUIIP). Whilst this mechanism of integrating a VAWG component onto a pre-existing programme offers a promising route to delivery at scale, the results highlight challenges with ensuring sufficient quality and VAWG expertise in VAWG component programmes when delivered by (largely) non VAWG specialists.[80]

Although there is no well-established minimum dosage for VAWG prevention interventions[81], our findings support emerging learning from the DFID What Works Programme that intensity of delivery is key to achieving results. In particular, this involves not only ensuring a sufficient programme duration (typically three+ years for social norms programmes), enough regular hours of delivery, effective mobilisation and maintaining high regular attendance.

As reported in part B, overall city level analysis suggests some heterogeneity with effects by cities mixed. It is plausible that many of the consistent city effects are in fact related to differences in implementation. In future designs of programmes, the recommendation is to more thoroughly unify high quality processes, delivery, and implementation, in order to hedge against variation if the goal is to examine overall (rather than area-specific) programme effects.

Achieving change in VAWG is a long-term process, requiring sustained resources and realistic timeframes

Part B reports a number of implementation weaknesses which likely contributed to the overall programme results including insufficient capacity coupled with limited resources for capacity building; insufficient design stage and no programme pilot; and significant delays to implementation which restricted overall programme duration. These weaknesses were largely the result of limited resources (time, financial and human). Had the SCI had sufficient resources to ensure sufficient capacity of key staff, pilot the approach, respond to emerging implementation challenges without shifting the underlying programme theory, and have the flexibility to sequence activities based on learning objectives, the programme may have achieved more positive results.

Programmes with longer term investment, sustained resources and realistic timeframes are likely to be more effective and to have both greater and more sustained impact.

***Who* delivers is just as important as *what* is delivered**

A key implementation weakness of the SCI was limited capacity of CSAs compounded by limited resources for capacity building and training. The findings highlight the importance of ensuring field staff have appropriate skills

[80] Note findings from the ICAI (2016) review of DFID's VAWG portfolio which found that VAWG component interventions tend to be lower than in VAWG-focused programmes, and weaker in their contextual analysis and their use of evidence about what works.

[81] A key research question currently being explored by the DFID What Works Programme is what intensity/dosage is needed for effective programming.

and capacity, are sufficiently motivated and committed and have the necessary support needed to model gender-responsive behaviours and motivate and support communities to tackle VAWG.

In addition to ensuring sufficient gender and VAWG expertise of partners particularly at programme design, we also draw recommendations from approaches such as SASA! which precede programme activities with a process by which the organisation supports its own staff to personally reflect on violence, power and relationships in their own lives, as a means for them to critical challenge these issues themselves and thus offer better support to programme beneficiaries.[82]

Importance of social norms theory for addressing VAWG

The findings highlight the pervasive nature of harmful social norms around IPV and VAWG in public spaces in the target areas. The findings suggest that the context of Madhya Pradesh descriptive norms – i.e. what people think other people do – may be a more powerful driver of VAWG than prescriptive norms – what people think other people expect them to do – with important implications for future programming. In this context, awareness raising about the prevalence of VAWG may exacerbate the very social norms the intervention intends to shift.[83]

Drawing on learning from the design of the SCI and emerging evidence and theory on tackling social norms in VAWG programming[84], programmes such as the SCI can be strengthened by applying a social norms approach and in particular would benefit from formative research to diagnose the specific norms underpinning IPV and public VAWG in the local context and subsequent activities designed to address them.[85]

Challenges of external validity of adapted programmes

The design of the SCI was based on an analysis of VAWG in Madhya Pradesh as well as evidence of global best practice in VAWG prevention. However, the results of the evaluation highlight the challenges of external validity when adapting elements of successful approaches and applying them in a different context. Without first piloting the different activities and adapting to the context, the design could have effectively lost some of the nuances and key aspects of approaches on which they were based.

Lessons from this evaluation show that innovative programmes of this nature which are based on evidence from other contexts, require a period of adaptation and learning to adapt approaches prior to implementation and scale up in order to ensure effectiveness of the approach and underlying theory.

Ineffective local VAWG response can significantly limits effectiveness of VAWG prevention

The findings demonstrate how an ineffective VAWG response can undermine prevention efforts. When systematic failures of police and legal systems exist, lack of trust in local response can present a significant barrier to reporting VAWG and prevent access to justice, whilst also playing an important part in sustaining local norms around impunity for perpetrators of VAWG.

It is important for prevention programmes to engage in strengthening VAWG response as a key strategy to ensure sustained social change. For programmes which engage in providing local response and mediation, women's safety must always remain the priority, and it is important to ensure that those who are mediating have sufficient skills and are provided with appropriate training and support.

[82] Raising voices, SASA! Approach in Uganda.

[83] Ibid. (2016)

[84] Alexander-Scott et al (2016)

[85] Note although the SCI was not designed with a Social norms approach, the Endline panel survey included social norms to track any changes in norms as well as attitudes and behaviour.

11. Insights for future VAWG evaluations

Our analysis has focused on what insights can be drawn for VAWG interventions. However, our findings also offer some important insights for future evaluations of VAWG programmes:

Importance of monitoring data to measure implementation fidelity.

Monitoring data is just as important as evaluation data and is key to tracking implementation fidelity. Evaluations of this kind would benefit from rigorous quantitative monitoring data that can be included in quantitative analyses at the level of the randomisation. Evaluations of this kind would benefit from more qualitative monitoring data including beneficiary feedback, observations of the quality of delivery, and monitoring of signs of backlash.

Robustness of VAWG measures

Measuring sensitive behaviours such as VAWG is notoriously challenging and a number of factors may effect a respondent's willingness to report from the way the question was asked, the manner of the enumerator, the presence of other in or around the household.

As part of the evaluation we were able to assess the reliability of commonly used VAW measures by both comparing consistency over time and by comparing direct measures with indirect measures. Both analyses suggest that the standard measures commonly used in evaluations of this form may not be reliable with important implications for future evaluations in this area. A key recommendation is that DFID take a strategic look across their portfolio at the robustness of measures around VAWG. The findings also highlight the advantages of including indirect measures alongside more direct measures of VAWG prevalence.

References

Alexander-Scott, M. Bell, E. Holden, J (2016) DFID Guidance Note: *Shifting Social Norms to Tackle Violence Against Women and Girls*. London: VAWG Helpdesk.

Brody, C, De Hoop T, Vojtkova, M, Warnock, R, Dunbar, M, Murthy, P and Dworkin, SL (2016) *Economic self-help group programmes for improving women's empowerment: a systematic review,* 3ie Grantee Final Review. London: International Initiative for Impact Evaluation (3ie).

Corstange, D. (2009). 'Sensitive questions, truthful answers? Modelling the list experiment with LISTIT.' *Political Analysis*. 17.1: 45-63.

DFID. (2011). 'Ethics, Principles for Research and Evaluation'.

DFID (2012) Guidance on Monitoring and Evaluation for Programming on VAWG. DFID Chase Team.
Ellsberg, M. and Heise, L. (2005). 'Researching Violence Against Women'. World Health Organization (WHO).

DFID. (2014). 'A Summary of the Evidence and Research Agenda for What Works: A Global Programme to Prevent Violence against Women and Girls'. (2014). *What Works: To Prevent Violence*. Department for International Development (DFID).

Fulu, E., Kerr-Wilson, A., and Lang, J. (2014). 'Evidence Review of interventions to prevent violence against women and girls'. *What Works to Prevent Violence Against Women and Girls*. DFID.

Heilman et al (2016) Whose Justice? Whose Alternative? Locating women's voice and agency in alternative dispute resolution responses to intimate partner violence. ICRW

Heise, L (2011) 'What Works to Prevent Partner Violence: An Evidence Overview' *What Works: To Prevent Violence Against Women and Girls.* DFID.

Holden, J (2015) *Working with men and boys: a rapid evidence review*. DFID VAWG Helpdesk Research Report no 77. London. DFID.

Indian Ministry of Health and Family Welfare (2005-06). 'National Family Health Survey'. (NFHS-3).

IPE/GHK (2015) Safe Cities Intitiative Project Closure Report. Non published report submitted to DFID

Jakimow T; Kilby P, 2006, 'Empowering women: A critique of the blueprint for self-help groups in India', Indian Journal of Gender Studies, vol. 13, no. 3, pp. 375 - 400, http://dx.doi.org/10.1177/097152150601300303

Jewkes, R, Flood, M and Lang, J (2014) 'From work with men and boys to changes of social norms and reduction in inequities in gender relations: a conceptual shift in prevention of violence against women and girls',*The Lancet violence against women and girls series 3*

Mackie G, Moneti F, Shakya H, Denny E (2015) What are social norms? How are they measured? *UNICEF/University of California, San Diego, Center for Global Justice*

Madhya Pradesh Urban Infrastructure Programme (2015) Safe Cities Initiative Project Closure Report. (unpublished report)

National Committee for Ethics in Social Science Research in Health (NCESSRH). (2003) 'Ethical Guidelines for Social Science Research in Health'. Section 4.2.8.

Neville S, Mclean Hilker L, Humphreys M, Husain S, Khan S, Lindsey S (2014) Madhya Pradesh Impact Evaluation of the Safe Cities Initiative Baseline Report, August 2014. Social Development Direct and Columbia University (unpublished)

Population council (2005) Building assets for safe, productive lives. NY, USA; Plan (2009) Because I am a girl: girls in the global economy. Plan International, Woking UK.

Raising Voices, LSHTM and CEDOVIP. (2015). Is Violence Against Women Preventable? Findings from the SASA! Study summarized for general audiences. Kampala, Uganda: Raising Voices.

Ricardo C, Eads M, and Barker G. (2011) *Engaging Boys and Men in the Prevention of sexual Violence.* Sexual Violence Inititative and Promundo. Pretoria, South Africa

Social Development Direct. (2011). Research Principles and Practices (unpublished)

Social Development Direct. (2011) Child Protection Policy (unpublished)

The Association of Social Anthropologists (ASA). (1999). 'Ethical Guidelines for Good Research'.

Vyas, S.; Watts, C. How does economic empowerment affect women's risk of intimate partner violence in low and middle income countries? A systematic review of published evidence**.** Journal of International Development (2009) 21 (5) 577-602. [DOI: 10.1002/jid.1500]

World Health Organisation (WHO). (2001). 'Putting Women First: Ethical and Safety Considerations for Research on Domestic Violence against women'. WHO.

WHO, LSHTM and the South African MRC (2013) *Global and regional estimates of violence against women: prevalence and health effects of intimate partner violence and non-partner violence.* Geneva: WHO

Women in Cities International (2012) Tackling gender exclusion. Experiences from the Gender Inclusive Cities Programme. WICI, Montreal.

Appendices

Appendix 1: List of experts consulted

The following table provides a list of attendees who took part in the Endline Design Workshop at the Habitat Centre, New Delhi on the 11th August 2011.

Name	Organisation
Arundhuti Roy Choudhury	SDA, DFID
Asmita Basu	MPUIIP
Suneeta Dhar	Jagori
Vimala Ramakrishnan	NCIS
Nandita Bhatla	ICRW
Shireen Jejeebhoy	Population Council
Suvakanta Roy Swain	FHI 360
Preet Rustagi	IHD, New Delhi
Indu Agnihotri	CWDS
Mamta Kohli	DFID

Appendix 2: Endline outcome measures

Descriptions of Outcome Variables

HL	VARIABLE.NAME	MEASURE	DEFINITION
H1a	INCOME_EARN	Women earning their own income	INCOME_EARN measures whether women are currently working for pay. It is coded as 1 if the respondent is currently working for pay (either in cash or in kind) and 0 if the respondent is not working or working without pay using YN1 in the endline survey instrument.
H1b	DECISIONS_ROLE	Role of women in household decision-making	DECISIONS_ROLE measures the role played by married or cohabitating women in household decision-making. It is coded as the number of household decisions, from a list of 4, over which married women have primary or joint decision-making power using EM1A - D in the endline survey instrument.
H1c	HHINCOME_CONTROL	Control exercised by women over household income	HHINCOME_CONTROL measures the control exercised by married or cohabiting women over household income. It is coded as the number of household spending decisions, from a list of 2, over which married women have primary or joint decision-making power using EM1A - B in the endline survey instrument.
H2a	MOBILITY	Mobility of women outside their slum	MOBILITY measures the mobility of women outside the colony during the day. It is coded as the frequency with which women travelled outside their home slum in the past 3 months on a scale from 0 (Never) to 5 (Everyday/Almost Everyday) using MS1 in the endline survey instrument.
H2b	SAFETY_DAY	Feelings of safety in public spaces during the day among women	SAFETY_DAY measures the feelings of safety in public spaces in and around the home slum during the day among women. It is coded as the number of public places, from a list of 11, that women feel safe or very safe visiting alone during the day using MS3 A-K in the endline survey instrument.
H2c	SAFETY_NIGHT	Feelings of safety in public spaces	SAFETY_NIGHT measures the mobility of women within their home slum after dark. It

		after dark among women	is coded as the number of public places, from a list of 11, that women feel safe or very safe visiting alone after dark using MS4 A-K in the endline survey instrument.
H3a	PIPV_ANY	Perpetration of any form of physical or sexual IPV by men	PIPV_ANY measures the perpetration of physical or sexual IPV against women by married or cohabiting men . It is coded as 1 if the respondent has perpetrated at least 1 of 7 forms of physical or sexual IPV in the last 3 months using PDV6-PDV12 in the endline survey instrument.
H3a	PIPV_SEV	Perpetration of a severe forms of physical IPV by men	PIPV_SEV measures the perpetration of severe forms ofl IPV against women by married or cohabiting men. It is coded as 1 if the respondent has perpetrated at least 1 of 3 forms of severe physical IPV in the last 3 months using PDV8-PDV10 in the endline survey instrument.
H3a	PIPV_SLAP	Perpetration of a specific form of IPV by men (directly reported)	PIPV_SLAP measures the perpetration of a specific form of IPV against women by married or cohabiting men, as reported directly in a survey question. It is coded as 1 if the respondent has slapped or thrown something at an intimate partner in the last 3 months using PDV6 in the endline survey instrument.
H3a	PIPV_LE	Perpetration of a specific form of IPV by men (list experiment)	PIPV_LE measures the perpetration of a specific form of IPV against women by married or cohabiting men, as reported indirectly in a list experiment. It is coded as the number of statements that a respondent indicates are true from a list of either 3 or 4 statements using LE8 and LE9 in the endline survey instrument.
H3b	EIPV_EMO	Experience of any form of emotional IPV by women	EIPV_EMO measures the prevalence of emotional abuse by an intimate partner among married or cohabiting women. It is coded as 1 if the respondent has experienced at least 1 of 6 forms of emotional abuse from an intimate partner in the last 3 months using EDV1-EDV6 in the endline survey instrument.

H3b	PIPV_EMO	Perpetration of emotional abuse by men	EIPV_EMO measures the perpetration of emotional abuse against women by married or cohabitating men. It is coded as 1 if the respondent has perpetrated at least 1 of 6 forms of emotional abuse in the last 3 months using PDV1A-PDV5 in the endline survey instrument.
H3c	EIPV_ANY	Experience of any form of physical or sexual IPV by women	EIPV_ANY measures the prevalence of physical or sexual IPV among married or cohabiting women . It is coded as 1 if the respondent has experienced at least 1 of 7 forms of physical or sexual IPV in the last 3 months using EDV7-EDV13 in the endline survey instrument.
H3c	EIPV_SLAP	Experience of a specific form of physical IPV by women (directly reported)	EIPV_SLAP measures the prevalence of a specific form of IPV among married or cohabiting women, as reported directly in a survey question. It is coded as 1 if the respondent was slapped or had something thrown at them by an intimate partner in the last 3 months using EDV7 in the endline survey instrument.
H3c	EIPV_LE	Experience of a specific form of physical IPV by women (list experiment)	EIPV_LE measures the prevalence of a specific form of IPV among married or cohabiting women, as reported indirectly in a list experiment. It is coded as the number of statements that a respondent indicates are true from a list of either 3 or 4 statements using LE6 and LE7 in the endline survey instrument.
H3d	EIPV_SEV	Experience of a severe physical form of IPV by women	EIPV_SEV measures the prevalence of severe forms of sexual IPV among married or cohabiting women . It is coded as 1 if the respondent has experienced at least 1 of 3 severe forms of physical IPV in the last 3 months using EDV9-EDV11 in the endline survey instrument.
H3d	EIPV_PERC	Perceived change in experience of IPV by women	EIPV_PERC measures the perceived change in experience of IPV among women who have been married/cohabiting for at least 2 years. It is coded as on a scale of 1 to 5, where 1 indicates experiencing much less and 5 indicates experiencing much more IPV at present than 2 years ago using EDV15 in the endline survey instrument.

H4a	PPV_ANY	Perpetration of any form of public violence or harassment against women/girls by men	PPV_ANY measures the perpetration of violence or harassment against girls/women in public spaces by men. It is coded as 1 if the respondent has perpetrated at least 1 of 5 forms of violence or harassment against a girl or woman in the last 3 months in a public space using PPV1-PPV5 in the endline survey instrument.
H4a	PPV_GROPE	Perpetration of a specific form of public violence or harassment against women/girls by men	PPV_GROPE measures the perpetration of a specific form of public harassment against girls/women by men, as reported directly in a survey question . It is coded as 1 if the respondent has touched, groped, stalked or flashed a girl/woman in a public place in the last 3 months using PPV2 in the endline survey instrument.
H4a	PPV_LE	Perpetration of a specific form of public violence or harassment against women/girls by men	PPV_LE measures the perpetration of a specific form of public harassment against women by men, as reported indirectly in a list experiment. It is coded as the number of statements that a respondent indicates are true from a list of either 3 or 4 statements using LE4 and LE5 in the endline survey instrument.
H4b	EPV_ANY	Experience of any form of public harassment or violence by women	EPV_ANY measures the prevalence of public violence and harassment among women . It is coded as 1 if the respondent has experienced at least 1 of 5 forms of violence or harassment in a public space in the last 3 months using EPV1-5 in the endline survey instrument.
H4b	EPV_GROPE	Experience of a specific form of public harassment or violence by women (directly reported)	EPV_GROPE measures the prevalence of a specific form of public harassment among women, as reported directly in a survey question. It is coded as 1 if the respondent has been touched, groped, stalked or flashed by a man in a public place in the last 3 months using EPV2 in the endline survey instrument.
H4b	EPV_LE	Experience of a specific form of public harassment or violence by women (list experiment)	EPV_LE measures the prevalence of a specific form of public harassment among women, as reported indirectly in a list experiment. It is coded as the number of statements that a respondent indicates are true from a list of either 3 or 4 statements

			using LE2 and LE3 in the endline survey instrument.
H4b	EPV_PERC	Perceived change in experience of public harassment or violence by women	EPV_PERC measures the perceived change in experience of public violence or harassment among women . It is coded as on a scale of 1 to 5, where 1 indicates experiencing much less and 5 indicates experiencing much more public violence or harassment at present than 2 years ago using EPV7 in the endline survey instrument.
H5a	RIPV_POLICE	Reporting IPV to the police or a protection officer	RIPV_POLICE measures the likelihood of reporting a hypothetical incident of hitting or choking by an intimate partner to the police, among married/cohabiting women. It is coded as on a scale of 1-4, where 1 is very unlikely to report and 4 is very likely to report using RDV9 in the endline survey instrument.
H5b	SIPV_SHG	Receiving support from SHG/SHG member after experiencing IPV	SIPV_SHG measures the expected level of helpfulness of an SHG/SHG member upon reporting an incident of IPV, among married/cohabiting women. It is coded as on a scale of 1-4, where 1 is very unhelpful and 4 is very helpful using RDV5 in the endline survey instrument.
H5b	SIPV_POLICE	Receiving support from police after experiencing IPV	SIPV_POLICE measures the expected level of helpfulness of the police upon reporting an incident of IPV, among married/cohabiting women. It is coded as on a scale of 1-4, where 1 is very unhelpful and 4 is very helpful using RDV10 in the endline survey instrument.
H5c	SPV_POLICE	Receiving support from police after experiencing public VAW	SPV_POLICE measures the expected level of helpfulness of the police upon reporting an incident of public harassment, among women. It is coded as on a scale of 1-4, where 1 is very unhelpful and 4 is very helpful using RPV2 in the endline survey instrument.
H5c	SPV_SHG	Receiving support from SHG/SHG member after experiencing public VAW	SPV_SHG measures the expected level of helpfulness of an SHG/SHG member upon reporting an incident of public harassment, among women. It is coded as on a scale of 1-4, where 1 is very unhelpful and 4 is very

			helpful using RPV4 in the endline survey instrument.
H5d	ACT_PV	Taking actions to prevent or respond to public VAW	ACT_PV measures the average likelihood of taking 3 hypothetical actions to prevent or respond to public harassment or violence against women. It is coded as on a scale of 1-4, where 1 is very unlikely and 4 is very likely using IN3A-B and IN3F in the endline survey instrument.
H5e	ACT_IPV	Taking actions to prevent or respond to IPV	ACT_IPV measures the average likelihood of taking 3 hypothetical actions to prevent or respond to IPV. It is coded as on a scale of 1-4, where 1 is very unlikely and 4 is very likely using IN3C-E in the endline survey instrument.
H5f	RPV_POLICE	Reporting public violence or harassment to a police or protection offier	RPV_POLICE measures the likelihood of reporting a hypothetical incident of stalking or groping in public place to the police, among women. It is coded as on a scale of 1-5, where 1 is very unlikely to report and 5 is very likely to report using RPV1 in the endline survey instrument.
H5g	IPV_STATE	Recognition of the state as an actor to bring about change in IPV	IPV_STATE measures the average importance of national, state and municipal governments for addressing IPV. It is coded as on a scale of 1-4, where 1 is completely irrelevant and 4 is very important using BH13A-C in the endline survey instrument.
H5g	PV_STATE	Recognition of the state as an actor to bring about change in public VAW	PV_STATE measures the average importance of national, state and municipal governments for addressing public VAW. It is coded as on a scale of 1-4, where 1 is completely irrelevant and 4 is very important using BH14A-C in the endline survey instrument.
H5j	TAKE_ENV	Willingness to engage others in the community on VAW	TAKE_ENV measures the willingness to engage others in the community on the issue of VAW. It is coded as 1 if the respondent accepts a set of petition-cards to distribute among community members using BH12 in the endline survey instrument.

H6a	GEMSCALE	Attitudes (Modified GEM Scale)	GEMSCALE measures gender equitable attitudes towards mobility, gender roles, household decionmaking, social relations and sexual relations. It is coded as the mean score of 5 attitudinal questions each measured on a scale from 1-5 where 1 indicates strong disagreement (agreement) and 5 indicates strong agreement (disagreement) with a gender equitable (inequitable) statement using AT1-5 in the endline survey instrument.
H6b	ATT_PIPV	Individual attitudes on IPV perpetration	ATT_PIPV measures the extent to which a respondent agrees that a man should beat his wife if she disobeys him. It is coded as on a scale of 1-5 where 1 is strongly disagree and 5 is strongly agree using SN15 in the endline survey instrument.
H6c	ATT_RIPV	Individual attitudes on reporting IPV	ATT_RIPV measures the extent to which a respondent agrees that a woman should report her husband to the police when he hits her. It is coded as on a scale of 1-5 where 1 is strongly disagree and 5 is strongly agree using SN17 in the endline survey instrument.
H6d	ATT_PPV	Individual attitudes on public harassment perpetration	ATT_PPV measures the extent to which a respondent agrees that sexually harassing women is harmless fun. It is coded as on a scale of 1-5 where 1 is strongly disagree and 5 is strongly agree using SN16 in the endline survey instrument.
H6e	ATT_RPV	Individual attitudes on reporting public harassment	ATT_RPV measures the extent to which a respondent agrees that a woman should report to the police when she experiences sexual harassment. It is coded as on a scale of 1-5 where 1 is strongly disagree and 5 is strongly agree using SN18 in the endline survey instrument.
H7a	DNORM_PIPV	Descriptive norms on IPV perpetration	DNORM_PIPV measures the perceived prevalence of IPV perpetration within a respondent's social network. It is coded as the number (out of 5 married men) that a respondent believes would beat their wives using SN3 in the endline survey instrument.

H7b	DNORM_RIPV	Descriptive norms on reporting IPV	DNORM_RIPV measures the perceived prevalence of IPV reporting within a respondent's social network. It is coded as the number (out of 5 married women) that a respondent believes would report their to the police if their husband hit them using SN9 in the endline survey instrument.
H7c	DNORM_PPV	Descriptive norms on public harassment perpetration	DNORM_PPV measures the perceived prevalence of public sexual harassment within a respondent's social network. It is coded as the number (out of 5 married men) that a respondent believes would harass women on the street for fun using SN6 in the endline survey instrument.
H7d	DNORM_RPV	Descriptive norms on reporting public harassment	DNORM_RPV measures the perceived prevalence of public sexual harassment reporting within a respondent's social network. It is coded as the number (out of 5 married women) that a respondent believes would report to the police if they experienced public sexual harassment using SN12 in the endline survey instrument.
H8a	PNORM_PIPV_M	Prescriptive norms on IPV perpetration (Men)	PNORM_PIPV_M measures the perceived approval of IPV perpetration within a respondent's social network of men. It is coded as a 1 if a respondent thinks that most out of 5 married men would approve of a man beating his wife if she disobeyed him using SN4A in the endline survey instrument.
H8a	PNORM_PIPV_F	Prescriptive norms on IPV perpetration (Women)	PNORM_PIPV_F measures the perceived approval of IPV perpetration within a respondent's social network of women. It is coded as 1 if a respondent thinks that most out of 5 married women would approve of a man beating his wife if she disobeyed him using SN5A in the endline survey instrument.
H8b	PNORM_RIPV_M	Prescriptive norms on reporting IPV (Men)	PNORM_RIPV_M measures the perceived approval of IPV reporting within a respondent's social network of men. It is coded as 1 if a respondent thinks that most out of 5 married men would approve of a woman reporting her husband to the police

			if he hits her using SN10A in the endline survey instrument.
H8b	PNORM_RIPV_F	Prescriptive norms on reporting IPV (Women)	PNORM_RIPV_F measures the perceived approval of IPV reporting within a respondent's social network of women. It is coded as 1 if a respondent thinks that most out of 5 married women would approve of a woman reporting her husband to the police if he hits her using SN11A in the endline survey instrument.
H8c	PNORM_PPV_M	Prescriptive norms on public harassment perpetration (Men)	PNORM_PPV_M measures the perceived approval of public sexual harassment perpetration within a respondent's social network of men. It is coded as 1 if a respondent thinks that most out of 5 married men would approve of a man sexually harassing a woman on the street using SN7 in the endline survey instrument.
H8c	PNORM_PPV_F	Prescriptive norms on public harassment perpetration (Women)	PNORM_PPV_F measures the perceived approval of public sexual harassment perpetration within a respondent's social network of women. It is coded as 1 if a respondent thinks that most out of 5 married women would approve of a man sexually harassing a woman on the street using SN8 in the endline survey instrument.
H8d	PNORM_RPV_M	Prescriptive norms on reporting public harassment (Men)	PNORM_RPV_M measures the perceived approval of public sexual harassment reporting within a respondent's social network of men. It is coded as 1 if a respondent thinks that most of out of 5 married men would approve of a woman reporting to the police if she experienced public sexual harassment using SN13 in the endline survey instrument.
H8d	PNORM_RPV_F	Prescriptive norms on reporting public harassment (Women)	PNORM_RPV_F measures the perceived approval of public sexual harassment reporting within a respondent's social network of women. It is coded as 1 if a respondent thinks that most of out of 5 married women would approve of a woman reporting to the police if she experienced public sexual harassment using SN14 in the endline survey instrument.

H5h	BH_ICARD	Individual support for Government VAW initiatives (Ticked or Signed Card)	BH_ICARD measures an individual's support for government initiatives on VAW through expressing support for a petition to the ULB on such an initiative on VAW. This measure is coded as a 1 if an individual ticked or signed a petition card in support and 0 otherwise using the endline behavioral dataset.
H5h	BH_FCARD	Garnering support for Government VAW initiatives (Signed Color Cards by Women)	BH_FCARD measures an individual's willingness and ability to garner women's support for a petiton on government VAW initiatives. This measure is coded as the number out of 5 cards on IPV or public VAW (depending on the dictionary assignment to the behavioral variation) that the respondent returns with signatures of women in their community expressing support for the initiative using the endline behavioral dataset.
H5h	BH_MCARD	Garnering support for Government VAW initiatives (Signed Color Cards by Men)	BH_MCARD measures an individual's willingness and ability to garnermen's support for a petiton on government VAW initiatives. This measure is coded as the number out of 5 cards on IPV or public VAW (depending on the dictionary assignment to the behavioral variation) that the respondent returns with signatures of men in their community expressing support for the initiative.
H5h	BH_ICARD_IPV	Individual support for government initiatives addressing IPV	BH_ICARD_IPV measures an individual's support for government initiatives related to IPV through an expression of support for a petition to the ULB. This measure is coded as a 1 if an individual ticked or signed a petition card (depending on the dictionary assignment to a behavioral variation) and 0 otherwise using the endline behavioral dataset.
H5h	BH_ICARD_PV	Individual support for government initiatives addressing Public VAW	BH_ICARD_PV measures an individual's support for government initiatives related to public VAW through expression of support for a petition to the ULB. This measure is coded as a 1 if an individual ticked or signed a petition card (depending on the dictionary assignment to a behavioral

			variation) and 0 otherwise using the endline behavioral dataset.
H5k	BH_FCARD_IPV	Garnering women's support on government initiatives addressing IPV	BH_FCARD_IPV measures an individual's willingness and ability to garner women's support for a petiton for ULB action on IPV . This measure is coded as the number out of 5 cards on IPV that the respondent returns with signatures of women in their community expressing support for the initiative using the endline behavioral dataset.
H5k	BH_FCARD_PV	Garnering women's support on government initiatives addressing Public VAW	BH_FCARD_PV measures an individual's willingness and ability to garner women's support for a petiton for ULB action on Public VAW. This measure is coded as the number out of 5 cards on Public VAW that the respondent returns with signatures of women in their community expressing support for the initiative.
H5k	BH_MCARD_IPV	Garnering men's support on government initiatives addressing IPV	BH_MCARD_IPV measures an individual's willingness and ability to garner men's support for a petiton for ULB action on IPV . This measure is coded as the number out of 5 cards on IPV that the respondent returns with signatures of men in their community expressing support for the initiative using the endline behavioral dataset.
H5k	BH_MCARD_PV	Garnering men's support on government initiatives addressing Public VAW	BH_MCARD_PV measures an individual's willingness and ability to garner men's support for a petiton for ULB action on Public VAW. This measure is coded as the number out of 5 cards on Public VAW that the respondent returns with signatures of men in their community expressing support for the initiative using the endline behavioral dataset.

Appendix 3: Randomisation of slum allocation: method and timing

The 250 slums included in the evaluation were pre-selected because they had been previously identified for inclusion in the larger MPUIIP programme, of which the SCI is just a small part[86]. The IP then selected two SHGs in each of the 250 slums, and identified men and boys who were invited to form one boys'/men's group in each slum. These 250 slums (and therefore the groups located within them) were then randomly assigned to the six treatment arms. This approach provides particular strength in terms of internal validity. External validity would have been enhanced somewhat if these 250 slums were themselves randomly sampled from a larger population of slums, but this was not possible given that the MPUIIP was already underway. The level at which random assignment took place is summarised in table 22.

Table 22: Selection and random assignment

Cities	Slums	SHGs	Boys'/men's groups
Four cities were selected by DFID and GoMP for inclusion in the Safe Cities Initiative.	250 slums were pre-selected as part of the broader MPUIIP. These slums were then randomly assigned to the six treatment arms.	Two SHGs were selected by GoMP and the IP in each of the 250 slums[87].	Boys and men were invited by the IP to form one new boys'/men's group in each of the 250 slums, through a mix of community meetings and conversations with individual men and boys[88].

Using a random assignment scheme meant there was no way to predict in advance whether or not any particular slum would receive a given type of intervention. However, the randomisation was set up using a blocking approach to make sure that the *distribution* of treatment and control slums was *balanced* in various ways. The randomisation took place *after* the baseline data collection. The most important dimensions of balance were:

- By city: each of the intervention combinations was distributed approximately equally across cities
- By slum characteristics, including:
  (a) Baseline levels of VAW (experience of IPV and experience of violence and harassment in public spaces);
  (b) Size;
  (c) Type of women's groups for the SHG treatments - whether they were pre-existing SHGs, microfinance groups or kitty groups;
  (d) Involvement in the previous Madhya Pradesh Urban Services Programme (MPUSP);
  (e) Below Poverty Line (BPL) score.

The randomisation procedure was conducted in the following two stages[89]:

[86] The 250 MPUIIP slums had been selected as those which would not be affected by relocation or redevelopment under the Rajiv Awas Yojna (RAY) National Housing Scheme for the Urban Poor.

[87] Where more than two SHGs existed in a slum, the two strongest SHGs were selected. Where no SHGs were present, kitty groups or MFIGs were selected instead. The strength of these groups varied considerably, with some not operational at the beginning of the intervention.

[88] In some cases, men and boys who were on the original list of direct beneficiaries were not contactable during the baseline survey. In these cases, the field research team identified additional boys or young men to take part in the evaluation. The IP then added these boys and men to the list of direct beneficiaries to participate in the Programme.

[89] Random assignment was implemented by a computer using an R script with a fixed seed to allow replication.

- In the first stage, blocks with six slums each were created in each city that were as similar as possible across dimensions (a) to (e) above (where 'similarity' was defined using Mahalanobis distance);[90]
- In the second stage, the six slums within each set of matched slums were assigned to one of the six treatment arms.

The randomisation took place after the baseline data collection and analysis for two reasons. First, the most important variables for blocking were the slum level prevalence of IPV and violence against women in public spaces. Such data was not previously available and was collected as part of the baseline survey. Second, it was important that baseline findings were not affected by the assignment and in particular that all respondents were in a similar position vis-a-vis the Programme at the time of the baseline research. This avoided the possibility that respondents in treatment slums would respond differently to those in control slums simply because they had been selected for treatment.

[90] Integer issues were dealt with by first randomly sampling 60 slums in each city for which the blocking procedure was implemented; the remaining slums were randomly allocated to treatment directly, ensuring that all slums had equal probabilities of assignment to treatment in each city. Mahalanobis distance measures the distance between two points in the covariate space taking account of the covariance between measures.

Appendix 4: Discarded designs

Table 23 outlines the main alternative designs for the RCT which were considered and maps out their benefits, as well as the reasons why each of these were rejected in favour of the selected design.

Table 23: Alternative designs considered for the RCT

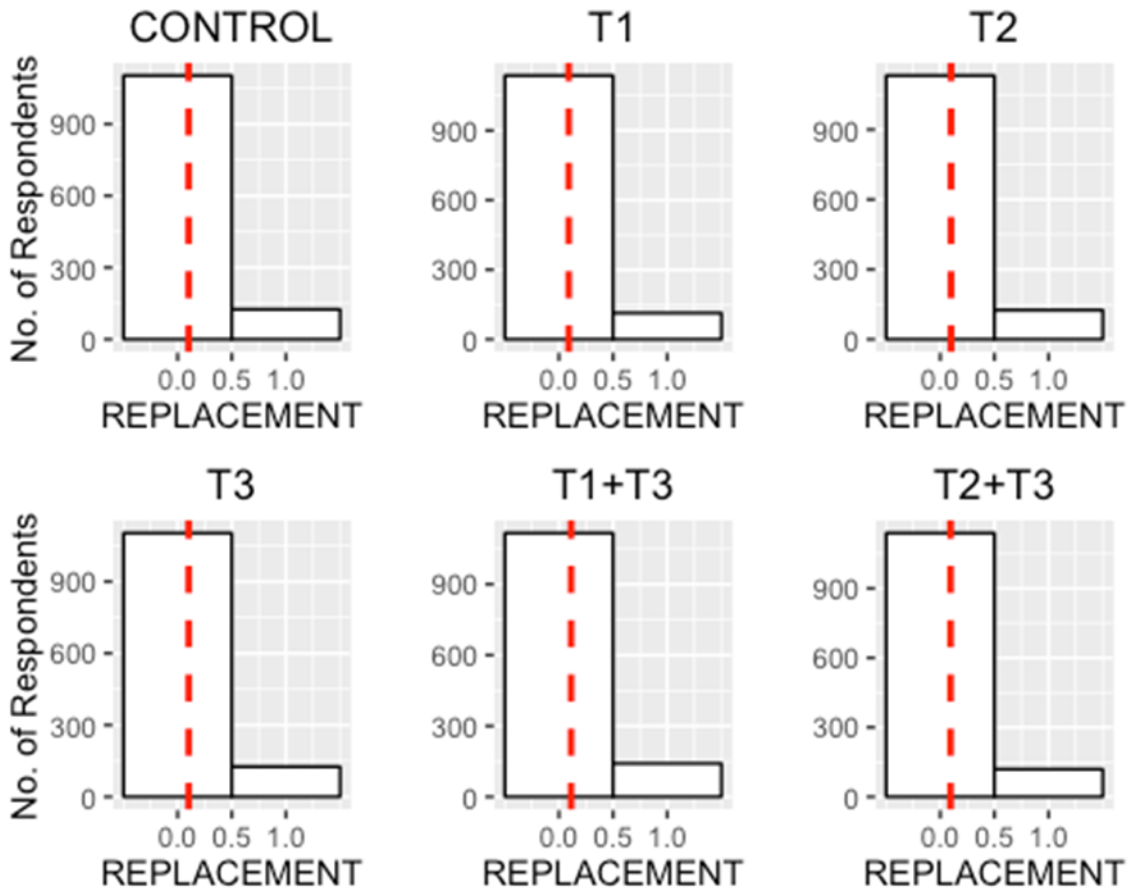
Alternatives considered	Benefit	Reason for rejection
1. Different randomisation schemes		
Use the SHG as the unit of randomisation.	Increase in statistical power and improved ability to assess spillover effects.	The men and boys groups would not be directly linked to the SHGs, so the randomisation would only be relevant for the women's interventions. Concern that within slum targeting would be imperfect and spillovers within slums would be too great.
Use "slum pockets" containing an SHG and a boys' sports club as a unit of randomisation.	As above.	The SHGs and men and boys' groups would not necessarily be geographically close together within the slum and group members could travel from anywhere in the slum to participate in them. The actual location of the group meeting within the slum is therefore less relevant. Identifying pre-existing groups appeared difficult.
Vary the intensity of the interventions (e.g. either one or two SHGs per slum).	Ability to draw controls from within treatment slums in order to allow for the treatment of all 250 slums, while still allowing for the assessment of spillover effects.	Importance of a pure control group to estimate total effects.
Use a three or four arm designs rather than factorial design.	Greater freedom to tailor an integrated intervention (rather than just combining separate interventions)	Weaker statistical power.
2. Identifying indirect beneficiaries in different ways		
As individuals within the immediate geographic vicinity ("slum pocket") of the directly treated population.	Potential to capture effects on a more specific indirect beneficiary group, rather than the slum population generally.	As noted above, group members could travel from any part of the slum to participate and the geography of the actual meeting is likely to be less relevant.
Individuals within the families or social networks of the directly treated population.	Ability to capture effects of direct beneficiaries' exposure to the initiative on their immediate family members, neighbours etc, based on a theory of change where the transmission channels are through these existing relationships.	Concern about the safety of participants in the evaluation if more than one household member was interviewed. In addition, the theory of change for the Safe cities Initiative is based on work with groups being an entry point to wider community engagement and dialogue. As such the initiative involves many

		activities at community level which are targeted at the general slum population.
3. Generating the boys' and men's groups through different processes		
Forming them using "seeds" from the baseline survey.	Groups would be representative of the population.	Artificial groups may not have the properties needed for an effective intervention.
Relying on pre-existing sports groups.	Potential to build on established groups in order to maximise programme effects and reduce risk of attrition.	Decision made by the implementation team: sports – or other groups – were unlikely to exist across the 250 slums, so the creation of new groups was the only viable option that would ensure consistency.
Creating them from volunteers from the network of women in SHGs.	Creating a direct link between the SHGs and men and boys' groups, building on existing relationships in order to potentially maximise programme effects.	Concern about the safety of participants in the evaluation if more than one household member was interviewed.

Appendix 5: Survey attrition

Survey Attrition

Overall, the endline survey had an attrition rate of **10.01**% i.e. **10.01**% of original baseline respondents could not be reached. The breakdown of reasons recorded for attrition are as follows. Importantly, we find that attrition rates are balanced across treatment groups. Plots below show the number of original vs. number of replacements by treatment group; the red dashed line denotes the mean attrition rate in the group.



We also regress attrition status on the treatment indicators; the p-value from an F-test of joint significance is 0.5702, which means we cannot reject the null hypotheses that all slopes are zero. This suggests that attrition is un-associated with treatment status.

Replacement in Case of Attrition

In the case an interview could not be conducted with the original respondent (i.e. an original respondent attrited from the survey sample), a replacement respondent was selected. The protocol for selection of replacement respondents was as follows:

Direct Beneficiaries

At the time of the baseline survey, the pool of direct beneficiary respondents was randomly drawn from a list of potential beneficiaries who had been identified by the IP. A list of replacement names for direct male

and female beneficiaries in each slum was generated using the "leftover" names in these lists, i.e. individuals whose names were provided in 2013 but were not randomly selected for an interview.

If a direct beneficiary who was interviewed in 2013 could not be re-interviewed at endline, a name was drawn from this replacement list. Where no extra names were available in a slum, no replacement interview was conducted.

Indirect Beneficiaries

If an indirect beneficiary interviewed in 2013 could not be re-interviewed, enumerators visited the dwelling to the immediate right of the original respondent's dwelling, confirmed that no interviews had been conducted in this dwelling, listed the eligible individuals in the household available for a survey and selected one of the eligible individuals using a selection table for random within-household selection.

If the original respondents dwelling could not be located or verified, no replacement was made.

Implications for Analysis

Since we do not have baseline covariate information for replacement respondents, we impute covariate-means calculated at the level of the slum & relevant beneficiary type for all replacement respondents. In the main analysis, we include all replacement respondents and control for their replacement status.

Appendix 6: Qualitative measures at endline

The following qualitative measures will be explored through the qualitative endline data collection:

Table 24: Qualitative measures at endline

Hypotheses	Outcome area	Research questions	Qualitative measures at endline	Methods
H1-7	**Social norms around IPV**	**Impact:** Has the intervention shifted social norms around IPV? **Processes of change:** Does this differ to quant data on prevalence? **Obstacles/barriers to change:** Which harmful social norms persist, why? In what forms?	**Social norms measure 1: Empirical expectations** around IPV: • Perceptions that IPV is common in the community • Perception that most women would not report their husband to the police for IPV	FGDs with men/women DB. Exercise vignette on IPV and following discussion.
H1-7, H13	**Social norms around IPV**	**Impact:** Has the intervention shifted social norms around IPV? **Processes of change:** Does this differ to quant data on individual attitudes (triangulation). **Obstacles/barriers to change:** Which harmful social norms persist, why? In what forms?	**Social norms measure 2: Normative expectations** around IPV: • Community acceptance of IPV • Community acceptance of IPV as a family issue • Community acceptance of not reporting IPV	FGD exercise 2 'myths and truths about VAWG' with men/women DB. Group consensus (agreement/disagreement) with harmful social and gender norms which underpin VAWG. Through this exercise we will explore collective attitudes (social norms) of what is normal/accepted within the reference group (the SHG/men and boys group). We will be able to compare if different types of interventions have shifted

				normative expectations of the reference group.
H1-7	**Social norms around IPV**	**Impact:** Has the intervention shifted social norms around IPV?	• **Social norms measure 3: social sanctions/rewards** around IPV • Social sanctions for men not perpetrating IPV • Social rewards for men perpetrating IPV • Social sanctions for women reporting IPV • Social sanctions for women reporting IPV • Social rewards for women not reporting IPV	FGDs with men/women DB. Exercise 2 and 5.
H1-7	**Social norms around violence and harassment against women in public places**	**Impact:** Has the intervention shifted social norms around violence and harassment against women and girls in public spaces? **Processes of change:** Does this differ to quant data on prevalence? **Obstacles/barriers to change:** Which harmful social norms persist, why? In what forms?	**Social norms measure 4: Empirical expectations** around violence and harassment in public spaces: • Perceptions that violence and harassment against women in public is common in the community • Perception that most women would not report cases of violence and harassment to the police	FGDs with men/women DB. FGD exercise 4 vignette and following discussion.
H1-7, H13	**Social norms around violence and harassment against women in public places**	**Impact:** Has the intervention shifted social norms around violence and harassment against women and girls? **Processes of change:** Does this differ to quant data on	**Social norms measure 5: Normative expectations** around violence and harassment in public spaces: • Community acceptance of violence and harassment against women and girls in public places	FGD exercise 2 'myths and truths about VAWG' with men/women DB. Group consensus (agreement/disagreement) with harmful social and gender norms which underpin VAWG.

		individual attitudes (triangulation). **Obstacles/barriers to change:** Which harmful social norms persist, why? In what forms?	• Community acceptance of not reporting violence and harassment in public	Through this exercise we will explore collective attitudes (social norms) of what is normal/accepted within the reference group (the SHG/men and boys group). We will be able to compare if different types of interventions have shifted normative expectations of the reference group.
H1-7	**Social norms around violence and harassment against women in public places**	**Impact:** Has the intervention shifted social norms around IPV?	**Social norms measure 6: social sanctions/rewards** around violence and harassment in public spaces • Social sanctions for men not perpetrating violence and harassment • Social rewards for men perpetrating violence and harassment in public • Social sanctions for women reporting violence and harassment in public • Social rewards for women not reporting violence and harassment in public	FGDs with men/women DB. Exercise 2 and 5. Exploration of existence of social consequences.
H4-6, H14,15,16	**Action to prevent VAWG** **Action to support survivors of VAWG**	Impact Why change happens How change happens Obstacles and barriers	Awareness of VAWG support services Trust and confidence in VAWG support services. Anticipated consequences for reporting violence Knowledge and awareness of reporting mechanisms	FGDs with men/women DB – exercise 3, 4 and 5

			Recent initiatives in the slum to address violence against women	
H1-16	**All**	**Unintended consequences of interventions**	Unintended consequences (positive and negative)	FGDs with men/women DBs. Exercise 3
9,10,12	**Income, decision making power**	**Impact:** Has the intervention increased income levels of DB? Has the intervention increased autonomy and decision making power of DBs/ wives of DBs How change happens	Economic empowerment of DB women Increased decision making power of women in the household	FGDs with men/women DB exercise 3.
H4-6, 8	**Women's feeling of safety** **Women's mobility**	**Impact:** Has the intervention improved women's feelings of safety and mobility in the slum? **How change happens**	Feelings of safety in and around the home slum Mobility in and around the home slum.	FGDs with women DB.
H1-16	**All**	**Diffusion effect** **Processes of change**	Other community members impacted by SHG/men and boys groups activities	FGDs with men/women. Exercise 3.
			Recommendations for how intervention could be improved Relevance of intervention	FGDs with men/women DBs – exercises 3 and 6

Appendix 7: Key informant interviews

The following table provides details of the key informants interviewed at endline.

Name	Post	City
Mohan Sarvan	Police Inspector, Bhopal Police Department	Bhopal
Renuka Mehta	Counsellor, Women's Police Department	Bhopal
Rani Tiwari	CDO	Bhopal
Rita Tuli	Counsellor, Women's Police Department	Bhopal
Sanjay Kumar	ULB Nodal Officer	Bhopal
Rahim Chauhan	CDO	Gwalior
Sonika Sharma	CDO	Gwalior
Santoshi Gupta	City Community Mobilisation Expert, City Cluster Team	Gwalior
Anita Mishra	T.I, Mahila Thana, Padav, Gwalior	Gwalior
Abhay Rajangaonkar	Additional Commissioner, (Municipal Corporation) Nodal Officer, MPUIIP, Gwalior	Gwalior
Rakesh Mahor	Counsellor, Ward – 37 & Sabhapati (Mayor in council)	Gwalior
Lokendra Jadhav	Community support agency (CSE) team member	Indore
Sadhna Paranjpe	Women Help Desk Counsellor/CWC Head and Head, Kasturba Gandhi Memorial Trust	Indore
KC Pandey	Protection officer	Indore
Rajendrda Mandloi	CDPO/ Protection officer	Indore
Vandna Yadav	Ward Councillor (ULB - Elected Representation), Choudhary Park Ward (51), Indore	Indore
Waseem Iqbal	GHK team member	Indore
Rahul Singh	Nodal Officer	Jabalpur
Rukmani Rajpoot	Counsellor at Family Counselling Centre	Jabalpur
Ajay Yadav	Community Development Officer	Jabalpur
Laxmi Yadav	Police Inspector, Women's Police Cell	Jabalpur

Appendix 8: Ethical guidelines for the evaluation

Table 25 below outlines the ethical guidelines which were developed during the inception phase for the evaluation, and which have been used by all team members to guide delivery of the baseline.

Table 25: Ethical guidelines for the evaluation of the Safe Cities Initiative (see following page)

Sources of suggested practice	Our approach
WHO (2001) Putting Women First: Ethical and Safety Considerations for Research on Domestic Violence against women – pages 10, 12-13. 22-23 http://whqlibdoc.who.int/hq/2001/WHO_FCH_GWH_01.1.pdf [India] National Committee for Ethics in Social Science Research in Health (NCESSRH) (2003) Ethical Guidelines for Social Science Research in Health - sections 4.1 http://www.fabtp.com/wp-content/uploads/2010/07/NCESSRH-Guidelines.pdf *The Association of Social Anthropologists (ASA)* Ethical Guidelines for Good Research – section 1.2 on "Anticipating Harms" http://www.theasa.org/ethics/Ethical_guidelines.pdf Social Development Direct's "Child and Vulnerable Adults Protection Policy"	1. Research assistants (RAs) will be trained to put the safety of participants as the main priority for the evaluation, to treat them with respect and sensitivity, and to be acutely aware of the risks of participation in the survey and other data collection methods and possible retaliatory action by perpetrators. 2. The evaluation will be framed and presented to community members as a general "safety and welfare survey" where men/boys and women will be asked about their lives in their colonies. This will also enable respondents to explain the survey to others safely. 3. Survey questions about violence will be worded sensitively with interviewer scripts to carefully introduce sections about violence, forewarn the respondent about the nature of the questions and give them the opportunity to stop the interview, or not to answer these questions. 4. Only one woman will be interviewed from each household and we will not interview women and men/boys from the same household. 5. We will aim for the survey interview to be conducted in a private space. As this is a panel survey, we recognise the difficulty of people achieving privacy in small homes which are of poor quality. 6. RAs will be trained to terminate or change the subject of discussion if the interview is interrupted by anyone. They will also forewarn the respondent of this approach. 7. RAs will be trained to detect signs of distress or trauma and to pause or stop the interview or discussion and provide information on support services on where the respondent can get help. 8. All respondents will be given a information card with numbers of local support services (at city and where possible, ward or colony level) for women at risk and VAWG survivors. This card will also include a range of other services, so it will not arouse suspicion if seen by another family or community member. Researchers will point one main VAW service provider out on the list to respondents who are illiterate. The research team will brief all of the main VAW service providers on the list and ensure they are ready to be contacted as needed. 9. There will be a clear procedure for handling cases where an RA is concerned that a respondent or a child under 18 is at risk of serious harm. Following discussion and agreement with the young person themselves, the concern will be reported to the Field Supervisor (FS) and in turn to the Research Manager and National Consultant for a joint decision. We acknowledge that this communications chain involves several steps, but each case will be addressed urgently. The case will be reported to a local service provider (e.g. shelter or specialist NGO) where it is judged that this will reduce - not increase - harm to the respondent. This will be done after careful consideration. Judgments will be made on a case by case basis depending on the case and the availability of local service providers that can genuinely help those at risk, not endanger them further. We do not propose identifying professional counsellors to accompany the evaluation team, but will make contact with existing local service providers.

The Association of Social Anthropologists (ASA) Ethical Guidelines for Good Research – section 1.4 on "Negotiating Informed Consent" http://www.theasa.org/ethics/Ethical_guidelines.pdf WHO (2001) Putting Women First: Ethical and Safety Considerations for Research on Domestic Violence against women – page 12 http://whqlibdoc.who.int/hq/2001/WHO_FCH_GWH_01.1.pdf (India) National Committee for Ethics in Social Science Research in Health (NCESSRH) (2003) Ethical Guidelines for Social Science Research in Health - section 4.2 http://www.fabtp.com/wp-content/uploads/2010/07/NCESSRH-Guidelines.pdf	1. RAs will read out a consent form in Hindi and ask all participants to provide verbal consent to take part in the evaluation. Participants will not be asked to sign the consent form (in order to avoid any anxiety this might cause, particularly if the participant cannot read or read very well). A copy of the consent information read out to participants will NOT be left with them, in case this is read by others. The information provided to the participants will include: - The purpose of the evaluation - Identity / contacts for researchers and sponsor - Why the individual has been selected for participation - What participation in the evaluation will entail - Any risks or benefits of the evaluation - Provisions for privacy, confidentiality and anonymity and any limitations - Future use of information - Right not to participate and to withdraw at any point 2. In order to gain access to the household to conduct the survey, permission may also be required from other household members. This is likely to include husbands, fathers, mothers and mothers-in-law. Although formal consent will not be sought from these other family members, RAs will be provided with a script to use to explain the evaluation to other household members, as necessary, and will be briefed on asking them for permission to enter and conduct interviews in the household. 3. RAs will request additional consent in FGDs and KIIs for use of voice recorders.
(India) National Committee for Ethics in Social Science Research in Health (NCESSRH) (2003) Ethical Guidelines for Social Science Research in Health - section 4.2.8 http://www.fabtp.com/wp-content/uploads/2010/07/NCESSRH-Guidelines.pdf *"Informed consent in the case of research with children (below the age of fourteen years) should be sought from the parents/guardians as well as the children themselves. Where the parents/guardians consent to participate, and the children have declined, the rights of the children should be respected. The consent from parents/ guardians*	1. RAs will request informed verbal consent from 15-17 year olds themselves, taking care not to put children under any pressure to give this (this will include the information outlined in the previous section above). 2. We will not seek consent of parents or caregivers for the participation of respondents aged 15-17 years. However, as noted in the previous section, permission to enter the household and interview household members will also be sought from heads of household/ other household members as necessary. 3. We will ensure the format, accessibility and content of the information and consent form is appropriate for 15-17 year olds. 4. Additional explanations will be provided by RAs as necessary.

should be waived only in special cases such as child abuse."	
Social Development Direct's "Child and Vulnerable Adults Protection Policy"	1. All RAs will receive basic training in child protection. 2. As above, informed consent will be asked of all children and additional explanations and support provided as needed.
WHO (2001) Putting Women First: Ethical and Safety Considerations for Research on Domestic Violence against women – pages 17-18 http://whqlibdoc.who.int/hq/2001/WHO_FCH_GWH_01.1.pdf (India) National Committee for Ethics in Social Science Research in Health (NCESSRH) (2003) Ethical Guidelines for Social Science Research in Health - sections 3.4, 4.3 http://www.fabtp.com/wp-content/uploads/2010/07/NCESSRH-Guidelines.pdf *The Association of Social Anthropologists (ASA)* Ethical Guidelines for Good Research – section 1.5 http://www.theasa.org/ethics/Ethical_guidelines.pdf	1. No names, addresses or other details that could allow identification of respondents will be recorded in the completed survey or in FGD notes and transcripts. In surveys and notes all respondents will be distinguished by a unique identifier code. 2. In FGDs, Field Investigators (FI) will be instructed not to mention identifiable names during the discussion. 3. In order to respect confidentiality and minimise the risk of harm, field investigators will state at the beginning of FGDs that they are not asking participants to talk about personal experiences (unless it is a group of individuals, like SHG, who already know each other, and feel comfortable talking about personal experiences together). 4. Key informants are always harder to anonymise in studies of this kind as findings will generally need to be quoted with reference to the kind of informant. During KIIs, RAs will be instructed to carefully check which information may be attributed and which is off the record or anonymous. Where appropriate, names will be withheld or changed. 5. As with all data files, audio files will be kept securely (password-protected on a secure server, CDs or devices locked in a secure cabinet) and no record of the respondents' names or addresses will be kept with the file, but unique identified codes will be used. 6. A separate record will be kept which links individual respondents' names and contact details to their unique identifier code (this is needed due to the fact that these are panel surveys and respondents need to be contacted again for the endline research). This will be a password protected file, accessed only be the FS and Research Manager. Following each survey round, this record will only be kept by the Research Manager on a secure server and deleted from FS' computers. 7. All researchers will receive training and detailed instructions on the importance of maintaining confidentiality and these procedures for doing so. 8. No researchers will conduct interviews in their own community or with people they know personally. 9. All respondents will be informed about the principle and limits of confidentiality and those who request more information will be informed about these procedures. The limits to confidentiality will be clearly defined and explained to respondents. Confidentiality will only be broken in a case where the respondent or a child under the age of 18 is judged to be at risk of serious harm (see section on "do no harm" above). 10. Care will be taken in both analysis and write-up stages to ensure that informants are referred to by code number or in generic terms (e.g. women, aged 17, Bhopal).
WHO (2001) Putting Women First: Ethical and Safety Considerations for	1. All researchers will be carefully selected and screened for negative attitudes towards women and VAWG.

Research on Domestic Violence against women – pages 19-20 http://whqlibdoc.who.int/hq/2001/WHO_FCH_GWH_01.1.pdf (India) National Committee for Ethics in Social Science Research in Health (2003) Ethical Guidelines for Social Science Research in Health - section 3.3 http://www.fabtp.com/wp-content/uploads/2010/07/NCESSRH-Guidelines.pdf	2. All researchers will receive training in all of the ethical issues and procedures, this will include discussion around possible scenarios which the researchers might encounter to make the training more tangible. 3. All researchers will receive specialized training on women's rights, child protection, VAWG, which will provide a mechanism for them to overcome their own biases, fears and stereotypes regarding the status of women and girls and VAWG survivors. The training will also provide opportunity for them to address any personal experiences of abuse. 4. More researchers than we need will be trained so that we can select only those which demonstrate an understanding of the ethical issues and do not demonstrate biases or negative attitudes which suggest they may undermine data collection and risk doing harm to others.
WHO (2001) Putting Women First: Ethical and Safety Considerations for Research on Domestic Violence against women – page 10 http://whqlibdoc.who.int/hq/2001/WHO_FCH_GWH_01.1.pdf	1. Researchers will work in teams of 5 (1 FS and 4 FIs) with at least two researchers in close proximity at any one time. The FS will maintain a record sheet of the interview schedule with information about where each researcher is at any one time. 2. All researchers will carry mobile phones with the contact details of other team members and key numbers, such as local police. 3. Researchers' safety will take precedence over the research and researchers will be instructed to terminate interviews or other fieldwork if they feel their safety is at risk. 4. Researchers will be briefed on practices to support one another during the research process. 5. Research teams will hold debriefing sessions at least once a week specifically to discuss their interview experiences and emotional responses to the research. In a case of acute distress or trauma, the researcher will be advised to take a break, and/or terminate their role and/or be referred to a counselor.
Social Development Direct – research principles and practices	1. Research questions will be phrased in appropriate and accessible language and asked in Hindi. 2. Researchers will provide any clarifications or explanations needed by respondents, taking care, however, not to influence answers. 3. FGDs will all be same sex. 4. In FGDs, careful facilitation and a mix of methods will be used (e.g. discussion and participatory exercises) to enable participation of all participants. 5. Each FGD participant will be given appropriate compensation for any transport costs plus refreshments. 6. Short summary publications will be produced in Hindi to highlight findings following analysis of the endline data and production of the final report. Disseminating findings at baseline could affect reporting at endline and undermine the evaluation. We are currently considering the best way for these to be made available to respondents, for example through programme staff and community-based organisations. This will be discussed with DFID and the IP, including way which findings could be shared with those who are illiterate.

WHO (2001) Putting Women First: Ethical and Safety Considerations for Research on Domestic Violence against women – pages 14-16 http://whqlibdoc.who.int/hq/2001/WHO_FCH_GWH_01.1.pdf	1. The survey instrument will be based on best practice, tried-and tested questions about prevalence, frequency and severity of public and private violence (e.g. specific questions) to try to capture the levels of violence as accurately as possible and avoid under-reporting. 2. All data collection tools with be field-tested to check for errors, inappropriate questions and impacts on respondents. 3. Female researchers will interview female respondents as experience shows that the sex of the interviewer has an impact on survey responses and reporting of violence. 4. Researchers will be trained in techniques to interview respondents in a sensitive and respectful manner, and inspire trust, so under-reporting is less likely. 5. Data will be checked and cleaned on a daily basis to minimise potential errors. 6. Data will be collected from different sources (e.g. survey, FGDs, KIIs) to cross-check and triangulate data and increase the likelihood of making accurate assessments.
Social Development Direct – Research Principles and Practices (unpublished)	1. Ethical principles that have informed our approach to control groups are: - There should be no deception about their involvement in the evaluation - Participants should not experience any harm through inclusion in the control group - Participants should benefit from being part of the control group - The benefits from having a control group should outweigh risks. 2. Our approach therefore involves: - Recognising that all colonies will potentially receive benefits from activities at city and state level during the programme - Recognising that colonies in the control group could *potentially* benefit from the upscaling of interventions found to be effective by the evaluation; - Applying all ethical practices outlined in this document equally to treatment and control groups, including readiness to provide support and referral to other agencies. 3. Control and treatment groups will be randomly selected after baseline data collection, so all survey respondents interviewed at baseline will not know whether they will be in the treatment or control groups. They will all have the same chance of being in either treatment or control. This means that they will have provided consent knowing there is a chance they might directly benefit from the programme, but equally a chance they they may not benefit. 4. The measures that will be gathered through this evaluation are the same/ similar to those that are routinely gathered in other surveys, for example DHS survey where there is no intervention at all. We therefore do not plan any additional activities to protect respondents in control groups: for the purpose of this evaluation, they will be treated the same as those in the treatment groups. 5. Given that the evaluation team will primarily be engaging with the 'direct beneficiaries' in treatment and control who are taking part in this evaluation, we will have very limited contact with members of the SHGs and men and boys' groups more widely. As such, we are relying on the IP to explain the evaluation to this wider group, to explain that by taking part in the programme they are agreeing to participate in an RCT and to seek their consent for this. The IP has been fully briefed

	on the purpose of the evaluation and the methods to be used so they can provide this information to those participating in the programme.

Appendix 9: Guidance on reading core results tables

The core results tables presented in part B are organised by outcome variable and give a complete description of the estimated effects of all treatments on each outcome for all subgroups. From these tables (see illustrations below) readers can see the raw data, the estimated average effects for each of the three treatments—SHG strengthening (T1), VAW training (T2), boys and men's interventions (T3), and the estimated interactions between treatments.

The core tables first introduce the variable and say how it was measured and what values the measure can take. The introductory text also links the measure to the survey instrument so that exact wording can be assessed.

Next come bar charts used to present the raw data. These charts present the data broken down by treatment arm for different subgroups. The height of each bar shows the average level of the variable in a group and the whisker-like markers coming out above and below the top of the bars show the size of the standard errors of the estimates of the average. If these stretch far above and below the top of the bar that indicates a lot of uncertainty in the mean – reflecting both the underlying variation in responses and the number of observations we have---- if they are tight, that represents a lot of certainty.

The first chart in the illustration below presents the raw data on female direct beneficiaries --- these are the women who *could have been selected* to take part in T2 and T3, thus properly these are the women "potential beneficiaries." This chart has six bars each corresponding to treatment arms, or combinations of the three treatments.

- Group "Control" is the group of potential female participants that were in slums that did not have any of the three treatments implemented.
- The group marked "T1 only" is the group of potential female participants that were in slums that had only T1 – SHG strengthening—implemented. T2 was not implemented for these women. Men in their slum did not receive T3.
- The group marked "T2 only" is the group of potential female participants that were in slums that had only T2 – SHG strengthening plus VAW training—implemented. Men in their slum did not receive T3.
- The group marked "T3 only" is the group of potential female participants that were in slums that had only T3 – boys and men interventions—implemented. Thus these women did not take part in programs directly, though they could still have been indirectly affected by T3.
- The group marked T1+T3 are women in slums in which both T1 and T3 was implemented.
- The group marked T2+T3 are women in slums in which both T2 and T3 was implemented.

The second chart provides a breakdown in the same way, but with a focus on the potential indirect beneficiaries (women in the slums that were not drawn from potential direct beneficiaries).

You can read the main effects of treatments directly from these bar charts. Specifically:

- Consider the effect of introducing T1 to slums in which T3 was not introduced. This is called $b(T1 \mid T3 = 0)$. You can see $b(T1 \mid T3 = 0)$ by comparing the bars for T1 only and Control. If these are very different (and if the whiskers are small), this suggests a large effect.
- Now consider the effect of introducing T1 to slums in which T3 *was* introduced. This is called $b(T1 \mid T3 = 1)$. To see this effect you can compare the T1+T3 results to the T3 only results.
- Consider now the average effect of T1 in all places in which it was introduced. Call this simple $b(T1)$. *To see* $b(T1)$ you look at the average of the difference between T1 only and Control and between T1+T3 and T3 only. This is the average of $b(T1 \mid T3 = 0)$ and $b(T1 \mid T3 = 1)$.

- To see the interaction of T1 and T3, which we call *b*(*T1T3),* you look at the difference between the differences between T1 only and Control and between T1+T3 and T3 only. Or *b*(*T1 | T3 = 1*) - *b*(*T1 | T3 = 0*).
- Similarly for *b*(*T2 | T3 = 0*), *b*(*T2 | T3 = 1*), *b*(*T2*), and *b*(*T2T3*), and similarly also for *b*(*T3 | T1 = 1*), *b*(*T3 | T2 = 1*), *b*(*T3 | T1 = 0 & T2 = 0*), *b*(*T3*).

The tables that follow the graphs report estimates of the same treatment effects but do so using regression models that control for baseline levels of the outcome---or measures closely related to these---whenever these are available as well as enumerator fixed effects, block fixed effects, an indicator for whether an endline respondent replaced a baseline respondent, and an indicator for whether any covariate means were imputed for a respondent. These controls are added in order to generate tighter estimates of treatment effects.

In total these tables report nine key estimates of causal effects of interest for each respondent category. The small table below clarifies which effect is reported in which cell.

Illustration of Mapping from Tables to Effects

Beneficiary Grp.	Main Effects				Interactions		
		T1	T2	T3		T1*T3	T2*T3
Female Direct	Treatment Effect	b(T1)	b(T2)	b(T3)	Interaction Effect	b(T1T3)	b(T2T3)
	SE				SE		
	Mean				Mean		
					T1 or T2 Effect	b(T1|T3=0)	b(T2|T3=0)
					T3 Effect	b(T3|T1=0)	b(T3|T2=0)
	N				N		

Note: Each pair of table reports nine treatment effects. The three numbers in the "Treatment Effect" row give the average effects of the T1, T2 and T3; the wo numbers in the "Interaction Effect" row give the T1,T3 interactions and the T2, T3 interactions.

The blocks on the left for each group reports b(T1), b(T2), and b(T3) along with related statistics. An example of this block is given below:

Illustration of Left Side Table (Main effects)

Beneficiary Grp.	Main Effects			
		T1	T2	T3
Female Direct	Treatment Effect	-0.01	0.034	0.004
	SE	0.026	0.026	0.021
	Mean	0.404	0.384	0.39
	N	1996	1996	1996

Here the numbers in the Treatment Effect row show the estimated average effects for each treatment. This corresponds to b(T1), b(T2), and b(T3) above. The first number here, -0.01 for example is the average effect

of T1, which averages the effects of T1 in slums in which T3 was implemented and slums in which it was not. The "Mean" estimate, 0.404, is the average estimate, across slums and other treatments conditions, in which T1 was not implemented. This means that we estimate that T2 changed average outcomes from 0.404 to 0.384. The numbers in the T2 and T3 columns give the same average effects for these other treatments.

Directly under the estimated effects we report the standard error of the effects estimate. This reports how confident we are in these estimated effects. If the standard error is less than half the size of the estimated effects, then the estimated effect is significantly different from 0 at the 95% level (more accurately: if the coefficient is more than 1.96 times the standard error there is less than a 5% chance that we would see such a big effect if there were no true effect). In those cases we add two stars (**). If the effect is significant at the 90% level we add one star (*) and if significant at the 99% level we add three stars (***).

Finally we add the N which says how many units were used in this analysis.

In the table on the right we show the interaction effects, corresponding to *b*(*T1T3*) and *b*(*T2T3*). These estimates are drawn from regressions with the same components, except that the treatment variables in the regressions on the right have not been normalized in order to facilitate interpretation of interactions. An example of this portion of the table is given below.

Illustration of Right Side Table (Interactions)

Interactions		
	T1*T3	T2*T3
Interaction Effect	-0.11**	-0.114**
SE	0.051	0.054
Mean	0.369	0.369
T1 or T2 Effect	0.045	0.091
T3 Effect	0.079	0.079
N	1996	1996

The numbers in the interaction effect row give the estimated interaction effects. Recall that *b*(*T1T3*) tells us how much bigger the effect of T1 is in those places in which T3 is implemented compared to those cases in which it is not. *b*(*T1T3*) also tells us how much bigger the effect of T3 is in those places in which T1 is implemented compared to those in which it is not---that is, the interaction effect can be interpreted in two ways. Similarly for *b*(*T2T3*).

For example, in the table above the estimate is -0.11 for the interaction between T1 and T3. The rows that follow help in interpreting how big this interaction is as compared to three numbers: the mean of the control condition, the estimate of the effect of T1 alone, and the estimate of the effect of T3 alone.

The mean is simply the mean of the variable in the control condition (and thus is the same as presented in the barchart above). The next row shows the effect of treatment T1 or T2 in the case in which T3=0 (the first number is for T1, the second for T2). The T3 Effect term shows the effect of T3 when T1 or T2 is 0. Note that this number will always be the same in both columns. The N remains as before and the star markings continue to indicate the same *p*-value thresholds as before.

In this example, the estimated effect of T3 when T1=0 is 0.079, which is positive; the interaction pulls this down by 0.11 points for a small negative effect of T3 when T1=1. The estimated effect of T1 in the absence of T3 is 0.45, this is pulled down also when T3 is present for an overall small average negative effect of T1. Note that the coefficient estimates in the right side of the table helps to better understand the overall

average effects presented in the left side of the table. The estimate of T3 in the left side table is a small positive average effect. It is a combination of the effects of T3 when T1=0 and T2=0, when T1=1, and when T2=1 which can all be seen separately in the right side table.

For the list experiment we report the estimate of the average outcome in the control group which is given by the coefficient on a variable that indicates whether a respondent received the long list (labelled "LongList" or "LongList Effect" in the tables"). The estimate of the effect of a given treatment is given by the interaction of the long list variable and the treatment (labelled "Treatment*Long") the estimate of interaction effects between two treatments is given by the triple interaction between treatments and the long list variables (labelled "Interaction*Long").

Appendix 10: Definitions of correlates

Descriptions of Model Covariates

COVARIATE.NAME	MEASURE	DEFINTION
B_SCASTE	Scheduled Caste/Tribe/OBC	SCASTE measures caste status among all respondents. It is coded 1 if a respondent reports belonging to a scheduled caste, scheduled tribe or other backward class, and 0 otherwise using Q14 in the baseline survey.
B_RELIGION	Main Religious Group	RELIGION measures religious affiliation among all respondents. It is coded is a 1 if a respondent reports being Hindu, and 0 if the respondent reports being any other religion (Muslim, Christian, Sikh etc) using Q12 in the baseline survey.
B_HHSCORE	Household Wealth	HHSCORE is a wealth index constructed for all respondents using primary component analysis of dummy variables indicating house material (observed by enumerator) and the water supply type, toilet type and ownership of various assets as reported by respondents using QE4, Q9, Q10 and Q11 in the baseline survey.
B_DISABILITY	Disability	DISABILITY measures disability status among all respondents. It is coded a 1 if a respondent reports having a physical or mental disability or permanent health problem that stops them from performing normal daily activities, and 0 otherwise using Q15 in the baseline survey.
B_CHILDVIOLENCE	Childhood Exposure to Violence	CHILDVIOLENCE measures childhood exposure to violence among all respondents. It is coded a 1 if a respondent reports either witnessing physical or emotional violence inflicted by various family members upon each other, or personally experiencing physical violence inflicted by a family member, and 0 otherwise using Q54, Q55, Q56, Q57 and Q58 in the baseline survey.
B_FROMMP	From MP	FROMMP measures whether a respondent was born in the state they currently reside

		in for all respondents. It is coded a 1 if a respondent reports Madhya Pradesh as their state of birth, and 0 otherwise using Q13 in the baseline survey.
B_EDU	Level of Education	EDU measures the highest education level attained by a respondent and is coded as an ordinal variable ranging from from 0:No schooling to 10: Post-graduate using Q17 in the baseline survey.
B_AGE	Age in Years	AGE measures the age of a respondent as is coded as the age in years reported in Q3 of the baseline survey .
B_INCOME_EARNING	Working for Income	INCOME_EARNING measures the earning status of the respondent. It is coded a 1 if the respondent reports "working and paid in cash" or "working and paid in kind" to Q18 of the baseline survey, and 0 otherwise.
B_NUMUNIONS	Number of Unions	NUMUNIONS is a count measure of the number of times a respondent has married or cohabited with a partner. It is coded as the number reported by the respondent in Q6 of the baseline survey.
B_PREGNANT	Pregnant	PREGNANT measures pregnancy status among all female respondents. It is coded if the respondent reports being pregnant in QW2 in the baseline survey.
B_NUMCHILDREN	Number of Children	NUMCHILDREN is a count measure of the number of children ever born to a respondent among all female respondents. It is coded as the total number of living sons, living daughters, sons who have died and daughters who have died reported by the respondent in Q7 of the baseline survey.
B_ANY_SONS	Ever Given Birth to a Son	ANY_SONS is a binary measure of whether a female respondent has ever given birth to a son. It is coded as 1 if a respondent reports having 1 or more sons who are alive or have died in Q7 of the baseline survey.
B_AGEATMARRIAGE	Age at Marriage	AGEATMARRIAGE measures the age of a respondent when they first got married as

		the age in years reported in Q5 of the baseline survey.
B_DOWRY_NOTSATISFY	Dowry Paid and Not Satisfied	DOWRYNOTSATISFY is a measure of dowry payment and satisfaction among all currently married/cohabiting female respondents. It is coded as a 1 if the respondent reports that a dowry was paid to the current spouse's at the time of marriage and also reports that they were not at all or only somewhat satisfied with the dowry amount, and 0 otherwise using QW5 and QW6 of the baseline survey.
B_DOWRY_SATISFY	Dowry Paid and Satisfied	DOWRYSATISFY is a measure of dowry payment and satisfaction among all currently married/cohabiting female respondents. It is coded as a 1 if the respondent reports that a dowry was paid to the current spouse's at the time of marriage and also reports that they were very satisfied with the dowry amount, and 0 otherwise using QW5 and QW6 of the baseline survey.
B_SPOUSEEDU	Spouse's Education Level	SPOUSEEDU measures the education level of a respondent's spouse and is coded as an ordinal variable ranging from from 0:No schooling to 10: Post-graduate using Q43 in the baseline survey.
B_SPOUSE_ALCOHOL	Spouse's Alcohol Use	SPOUSE_ALCOHOL measures the frequency of alcohol consumption for a respondent's spouse among all currently married/cohabiting female respondents. It is coded as a 1 if a respondent reports that their spouse drank at least a few times a month, and 0 otherwise, using QW7 in the baseline survey.
B_NUCLEARFAM	Nuclear Family	NUCLEARFAM measures the family structure of a respondent among all currently married/cohabiting female respondents. It is coded a 1 if a respondent does not report any persons besides her spouse and children living in the household on a permanent basis, and 0 otherwise using Q8 in the baseline survey.

B_SPOUSEWORKING	Spouse's Work Status	SPOUSEWORKING measures the earning status of the respondent's spouse among currently married/cohabiting women. It is coded a 1 if the respondent reports that her spouse is "working and paid in cash" or "working and paid in kind" using Q45 of the baseline survey, and 0 otherwise.
B_HUSB_AGEDIFF	Spousal Age Difference	HUSB_AGEDIFF is a measure of how much older a respondent's current spouse is among currently married and cohabiting women. It is coded as the difference in years between the respondent's reported age and the reported age of her spouse using Q3 and Q41 in the baseline survey.
B_ACCESSSANITATION	Access to Sanitation	ACCESSSANITATION is a measure of access to water supply and toilet in the respondent's own home. It is coded as a 0 if the respondent does not report having access to his/her own water source or his/her own toilet, 1 if they report having access to either and 2 if they report having access to both using Q9 and Q10 of the baseline survey.
B_MARITALSTAT	Married or Cohabitating	MARITALSTAT is a measure of the respondent's current marital status. It is coded as a 1 if the respondent reports being currently married or cohabitating with a partner and 0 otherwise using Q4 in the baseline survey.
B_MALE_ALCOHOL	Male Alcohol Use	MALE_ALCOHOL measures the frequency of alcohol consumption among men in the last 3 months as reported by male respondents. It is coded 1 if a respondent reports consuming a drink containing alcohol at least a few times a month and 0 otherwise using QM3 in the baseline survey instrument.
B_VAWATT	Male Attitudes on VAW	VAWATT measures male attitudes towards public VAW at the individual level. It is coded as the number of responses to Q75, 76, 78, 79 and 80 (statements blaming women for public violence against women) that a respondent agrees or partially agrees with in the baseline survey.

B_SLUM_ALCOHOL	Slum Male Alcohol Use	SLUM_ALCOHOL is a measure of the average frequency of alcohol consumption among males in the slum. It is calculated for each respondent as the mean of responses to QM3 (coded 1 if a respondent reports consuming alcohol at least a few times a month and 0 otherwise) for males in the slum as reported in teh baseline survey. In the case of male respondents, their own response is excluded from the calculation of this number.
B_SLUM_PCTBPL	Slum Pct. Below Poverty Line	SLUM_PCTBPL is a measure of average poverty. It is coded for each respondent as the percentage of the slum population that lives below the poverty line, using data provided by GHK prior to the baseline survey.
B_SLUM_PROBDIFF1	Slum Fractionalization	SLUM_PROBDIFF1 is a measure of fractionalisation along religion, caste or migrant status. It is calculated for each respondent as the probability that he/she meets someone different from him/her in the slum on one of these 3 dimensions. Religion is coded 1 through 9 based on whether an individual reports being Hindu, Muslim, Buddhist, Sikh, Christian, Jain, Parsi, or having no religion. Caste is coded 1 through 4 based on whether an individual reports being from a Schedule Cate, Scheduled Tribe, Other Backwards Caste, or None of the above. , Migrant status is coded 1 if the respondent reports being born in MP and 0 otherwise. This variable coding draws from Q12, Q13 and Q14 in the baseline survey.
B_SLUM_SIZE	Slum Number of Households	SLUM_SIZE is a measure of slum size. It is coded for each respondent as the number of households in the slum, using data provided by GHK and collected by New Concept during the slum-mapping process for baseline data collection.
B_SLUM_MALEUNEMPLOY	Slum Male Unemployment	SLUM_MALEUNEMPLOY is a measure of the average level of male unemployment in the slum. It is calculated for each

		respondent as the mean of responses to Q18 for males in the slum as reported in the baseline survey. In the case of male respondents, their own response is excluded from the calculation of this number.
B_SLUM_VAWATT	Slum Male Attitudes on VAW	SLUM_VAWATT is a measure of the average male attitudes towards violence against women. It is calculated for each respondent as the mean number of responses to Q75, 76, 78, 79 and 80 (statements blaming women for public violence against women) that males in the slum agree or partially agree with as reported in the baseline survey. In the case of male respondents, their own response is excluded from the calculation of this number.
B_DECISIONS_ROLE_LAG	Women's role in household decision-making	DECISIONS_ROLE_LAG measures the role played by currently married or cohabiting female respondents in household decision-making. It is coded as the number of decisions, from a list of 4, for which the respondent is the joint or primary decision-maker using Q51_1, Q51_3, Q51_4, and Q51_5 in the baseline survey instrument.
B_HHINCOME_CONTROL_LAG	Women's role control of household income	HHINCOME_CONTROL_LAG measures the role played by currently married or cohabiting female respondents in household decision-making. It is coded as the total number of decisions on household purchases, from a list of 2, for which the respondent is the joint or primary decision-maker using Q51_3 and Q51_4 in the baseline survey instrument
B_MOBILITY_OUT	Women's mobility outside their home slum	MOBILITY_OUT measures the mobility of female respondents based on how often they travel outside their home slum. It is coded as the frequency with which a respondent travelled outside their home slum in the last month, on a scale of 0 to 5, 0 for respondents that did not leave their home slum and 5 for respondents that left

		on most days using Q62 in the baseline survey instrument.
B_SAFETY_DAY_LAG	Women's feelings of safety in their slum during the day	SAFETY_DAY_LAG measures how safe women feel in their slum during the day. It is coded as reported feelings of safety to work in the colony during the day measured on a scale from 1 to 4 with 1 coded as feeling very safe and 4 coded as feeling not safe at all using Q64_2 in the baseline survey instrument.
B_SAFETY_NIGHT_LAG	Women's feelings of safety in their slum at night	SAFETY_NIGHT_LAG measures how safe women feel in their slum at night. It is coded as the mean of reported feelings of safety measured on a scale from 1 to 4 with 1 coded as feeling very safe and 4 coded as feeling not safe at all to leave home in ther colony after dark using Q Q64_6 and Q64_7 in the baseline survey instrument.
B_VAW_PERP	Men's perpetration of harassment or VAW	VAW_PERP measures perpetration of harassment, physical violence, or sexual violence against women or girls in the home or public spaces, as reported by male respondents. It is coded 1 if the respondent answers yes to having perpetrated at least 1 of 15 forms of public harassment or violence against women in the last 12 months, and 0 otherwise using QM19, QM20, QM21, QM22, and QM23 in the survey baseline instrument.
B_VAW_PERPSEV_LAG	Men's perpetration of severe forms of IPV	VAW_PERPSEV_LAG measures perpetration of severe forms of physical violence against women or girls in the home or public spaces, as reported by male respondents. It is coded 1 if the respondent answers yes to having perpetrated at least 1 of 4 forms of violence against women in the last 12 months, and 0 otherwise using QM23_6, QM23_7, QM23_8, and QM23_9 in the baseline survey instrument.
B_PIPV_SLAP_LAG	Men's perpetration of a particular form of IPV	PIPV_SLAP_LAG measures perpetration of against women or girls in the home or public spaces, as reported by male respondents. It is coded 1 if the

		respondent answers yes to having perpetrated at least 1 of 2 forms of violence against women in the last 12 months, and 0 otherwise using QM23_03, and QM23_04 in the baseline survey instrument.
B_EMOTVAW_PREV	Women's experience of emotional abuse	EMOTVAW_PREV measures the experience of emotional abuse among currently married or cohabiting female respondents. It is coded 1 if the respondent answers yes to having experienced at least 1 of 2 forms of emotional abuse in the last 12 months, and 0 otherwise using QW30_2 A-B in the survey instrument.
B_PIPV_EMO_LAG	Men's perpetration of emotional abuse	PIPV_EMO_LAG measures perpetration of emotional abuse of women or girls in the home or public spaces, as reported by male respondents. It is coded 1 if the respondent answers yes to having perpetrated at least 1 of 2 forms of emotional abuse against women in the last 12 months, and 0 otherwise using QM23_01, and QM23_02 in the baseline survey instrument.
B_PHYSSEXVAW_PREV	Women's experience of physical or sexual IPV	PHYSSEXVAWPREV measures the prevalence of physical or sexual IPV among married or cohabiting women. It is coded as 1 if the respondent has experienced at least 1 of 9 forms of physical or sexual IPV in the last 12 months using QW30_04-QW30_12 in the baseline survey instrument.
B_EIPV_SLAP_LAG	Women's experience of a form of physical IPV	EIPV_SLAP_LAG measures the experience of a particular form of IPV among married or cohabiting women. It is coded as 1 if the respondent has experienced at least one of two forms of physical IPV in the last 12 months using and QW30_04 and QW30_05 of the baseline survey instrument
B_PHYSVAW_SEV	Women's experience of severe physical IPV	PHYSVAW_SEV measures the experience of severe forms of non-sexual physical IPV among currently married or cohabiting female respondents. It is coded 1 if the

		respondent answers yes to having experienced at least 1 of 3 SEVERE forms of non-sexual physical IPV in the last 12 months, and 0 otherwise using in the survey instrument.
B_PPV_GROPE_LAG	Men's perpetration of one type of harassment	PPV_GROPE_LAG measures perpetration of harassment against women or girls in the home or public spaces, as reported by male respondents. It is coded 1 if the respondent answers yes to having perpetrated at least 1 of 3 forms of harassment in the last 12 months, and 0 otherwise using QM20, QM21, and QM22 in the baseline survey instrument.
B_PUBLICVAW_PREV	Women's experience of harassment and public VAW	PUBLICVAW_PREV measures the experience of harassment, physical violence, and sexual violence against women in public spaces among all female respondents. It is coded 1 if the respondent answers yes to having experienced at least 1 of 12 forms of violence or harassment in public spaces in the last 12 months, and 0 otherwise using QW13 A-L in the baseline survey instrument.
B_EPV_GROPE_LAG	Experience of one type of public harassment	EPV_GROPE_LAG measures the experience of a particular form of public harassment among women. It is coded as 1 if the respondent has experienced a form of public harassment (groping) in the last 12 months using QW13_06 of the baseline survey instrument.
B_ACT_PV_LAG	Actions taken to prevent or respond to public VAW	ACT_PV_LAG measures engagement in preventive or responsive activity that addresses public harassment and violence against women, among all respondents. It is coded as 1 if the respondent reports taking at least 1 out of a list of 3 possible actions to prevent or respond to publc VAW in the past 12 months using q128_1,q128_2, and q128_6 in the baseline survey instrument.
B_ACT_IPV_LAG	Actions taken to prevent or respond to IPV	ACT_IPV_LAG measures engagement in preventive or responsive activity that addresses IPV, among all respondents. It is

		coded as 1 if the respondent reports taking at least 1 out of a list of 3 possible actions to prevent or respond to IPVin the past 12 months using q128_3, q128_4, and q128_5 in the baseline survey instrument.
B_ULBACT_LAG	State actions to prevent or respond to VAW	ULB_LAG measures state engagement in programs addressing VAW, among all respondents. It is coded as 1 if the respondent reports that the ULB has taken any 1 out of a list of 8 possible actions to prevent or respond to VAW in the past 12 months using Q119_1-Q119_8 in the baseline survey instrument.
B_ACTIONS_VAW	Actions taken to prevent or respond to VAW	ACTIONS_VAW measures engagement in preventive or responsive activity that addresses violence against women, among all respondents. It is coded as 1 if the respondent reports taking at least 1 out of a list of 7 possible actions to prevent or respond to VAW in the past 12 months using Q128 in the baseline survey instrument.
B_GEMSCALE_LAG	Gender Equitable Attitudes	GEMSCALE_LAG measures gender equitable attitudes towards mobility, gender roles, household decionmaking, social relations and sexual relations. It is coded on a scale from 1 to 4 as the mean level of agreement to attitudinal questions representing gender equitable attitudes using Q77, Q70, Q68, Q81, and Q83 in the baseline survey instrument.
B_ATT_PIPV_LAG	Attitudes towards perpetrating IPV	ATT_PIPV_LAG measures the extent to which a respondent agrees that there are times when a wife should be beaten. It is coded as on a scale of 1-4 where 1 is disagree and 4 is agree using Q73 in the baseline survey instrument.
B_ATT_RIPV_LAG	Attitudes towards reporting IPV	ATT_RIPV_LAG measures the extent to which a respondent agrees that domestic violence should be not be discussed outside the family. It is coded as on a scale of 1-4 where 1 is disagree and 4 is agree using Q89 in the baseline survey instrument.

B_ATT_PPV_LAG	Attitudes towards perpetrating public VAW	ATT_PPV_LAG measures the extent to which a respondent agrees that eve teasing is harmless fun. It is coded as on a scale of 1-4 where 1 is disagree and 4 is agree using Q79 in the baseline survey instrument.
B_ATT_RPV_LAG	Attitudes towards reporting public VAW	ATT_RIPV_LAG measures the extent to which a respondent agrees that women are at fault for public harassment against them. It is coded as on a scale of 1-4 where 1 is disagree and 4 is agree using Q75 in the baseline survey instrument.