ECWG
EVIDENCE CONSORTIUM ON WOMEN'S GROUPS

FEBRUARY 2021

Evaluating the effects of interventions with women's groups on health outcomes:

Resources for programme planners and evaluators

Sapna Desai[1] | Lu Gram[2] | Neha Kumar[3] | Audrey Prost[2]

1. Population Council, New Delhi
2. Institute for Global Health, University College London
3. International Food Policy Research Institute, Washington, DC

Acknowledgements

These guidelines draw from the work of colleagues at several institutions, including University College London, International Food Policy Research Institute and the Population Council. We are grateful to Dr Kalyani Raghunathan, IFPRI, Dr Thomas de Hoop, AIR and Dr Neeta Goel and Shubha Jayaram of the Bill & Melinda Gates Foundation for helpful review and feedback on these guidelines. We would like to thank co-members of the Evidence Consortium on Women's Groups, specifically Dr Leigh Anderson, Dr Gary Darmstadt and Roopal Jyoti Singh for their support. Finally, we thank Phil Esra, Jane Garwood and Sharon Wallace at the American Institutes for Research for excellent editorial inputs. The Evidence Consortium on Women's Groups is funded by a grant from the Bill & Melinda Gates Foundation.

Contents

Page

Introduction 1
Common evaluation challenges 1
1. Choosing a health outcome 2
Building a theory of change 2
Identifying health outcome measures 5
2. Measurement scope—group or population? 6
3. Minimising risk of bias 7
4. Process evaluation and intervention reporting 9
Conclusion 12
References 13

Figures

Figure 1. Theory of change for a community mobilisation intervention to address violence against women 3
Figure 2. Theory of change for self-help groups to improve nutrition outcomes in Bihar, India 4

Tables

Table 1. Population- versus member-focussed sampling 7
Table 2. Mitigating risk of bias in studies of group-based interventions 8
Table 3. Suggested checklist for reporting on group-based behaviour change and health interventions 11

Introduction

There is growing interest in the potential of women's groups to improve health. Large-scale government programs, such as India's National Health Mission and National Rural Livelihoods Mission, promote group-based interventions to improve maternal and child health. The evidence base has also expanded considerably in recent years, with a growing number of impact evaluations that examine the effects of group-based interventions on health outcomes.

Unfortunately, recent evidence syntheses have also identified several weaknesses in evaluations (Anderson et al., 2019; Gichuru et al., 2016; Kumar et al., 2018; Orton et al., 2016). Orton and colleagues' review of evaluations of microfinance-based groups found that 17 of 31 evaluations were of low quality, mostly due to selection bias. A systematic review of 44 studies in India found that one in three experimental and quasi-experimental studies were at high risk of bias due to selection bias, failure to pre-specify primary outcomes, and not accounting for missing data (Desai et al., 2020). Further, inconsistent measurement approaches and insufficient descriptions of group models make evidence syntheses difficult, which limits learning across contexts.

Public health researchers have provided guidance for intervention design, evaluation, and reporting, such as the Medical Research Council guidance on developing and evaluating complex interventions, the CONSORT guidelines for reporting on randomized control trials, and the TIDieR guidelines for describing interventions. This document aims to complement these tools. We provide guidance and resources specific to evaluations of the effect on health outcomes of interventions with women's groups—with a focus on design and reporting.

Common evaluation challenges

Drawing from evidence syntheses and our experience in conducting evaluations, we identified four common challenges to evaluating women's group interventions to improve health (Box 1) (Prost et al., 2013; Kumar et al., 2018; Gram et al., 2019; Desai et al., 2020). While two common approaches—using a theory of change and conducting process evaluations—are not restricted to women's group evaluations, this document addresses their specific application to women's groups. Addressing selection bias at the group and individual levels and identifying the evaluation's measurement scope are challenges faced in evaluating all group-based interventions.

Box 1. Challenges in evaluating the effects of interventions with women's groups on health outcomes

- Identifying health outcomes aligned with a theory of change
- Appropriate measurement scope—group or population-level outcomes
- Addressing selection bias at the group and member level
- Lack of intervention process monitoring and evaluation

This document aims to provide guidance and a collection of resources, structured in four parts:

1. Choosing health outcome measures
2. Identifying the appropriate scope of measurement
3. Minimising risk of bias
4. Process evaluation and reporting

We provide links to resources and examples from previous evaluations in boxes throughout the document, and conclude with a checklist to guide future evaluations.

1. Choosing a health outcome

Planners and evaluators typically initiate the evaluation design process by developing a theory of change. We offer suggestions on how to incorporate factors specific to women's groups at the evaluation design stage, followed by three criteria to help choose health outcomes.

Building a theory of change

Readers will be familiar with existing resources that describe how to develop a theory of change and how to adapt the principles to specific interventions (Davies, 2018; De Silva et al., 2014; White, 2009). Briefly, a theory of change helps to identify and measure indicators in a hypothesized causal chain, from intervention to outcome (Weiss, 1997; Davies, 2018). Theories of change are intended to be developed in collaboration between implementers, researchers, and other stakeholders—and modified throughout the intervention and evaluation process. They typically incorporate the intervention's context, long-term benefits, the process of change, and assumptions underlying each step, and use either existing evidence or a theoretical proposition to links steps in the causal chain (Vogel, 2012; White, 2009; De Silva, 2014).

Resource Box 1: Theory of change

A theory-driven approach to the MRC guidance on complex interventions, De Silva et al., 2014

Representing a theory of change, Davies et al., 2018

Theory-based impact evaluation, White, 2009

A theory of change specific to women's group interventions will incorporate assumptions and causal pathways related to the *type of women's group* and *how the group addresses health*. Women's groups vary widely in their organization, size, and implementation models. The theory of change should include assumptions related to group features such as the organising purpose (e.g., microfinance, livelihoods, or collective action for health), membership criteria, and meeting frequency.

There is considerable diversity in *how* interventions with women's groups aim to improve health (Gram et al., 2020). Approaches can include information dissemination to group members to address individual awareness while using the group as a platform to reach more women. Others may aim to improve members' skills and capacity to act as a group, building individual resources as well as collective action. Interventions aimed at community mobilisation seek to build capabilities through participatory methods to address direct and underlying determinants of health. Identifying the intervention's approach to working with groups can help incorporate assumptions while building the theory of change.

We recommend evaluators draw from these resources on women's groups to help identify assumptions related to the type of group and its approach to health. Figures 1 and 2 provide examples of theories of change for two different types of groups and intervention approaches.

Resource Box 2: Classifying women's groups and their approach to health

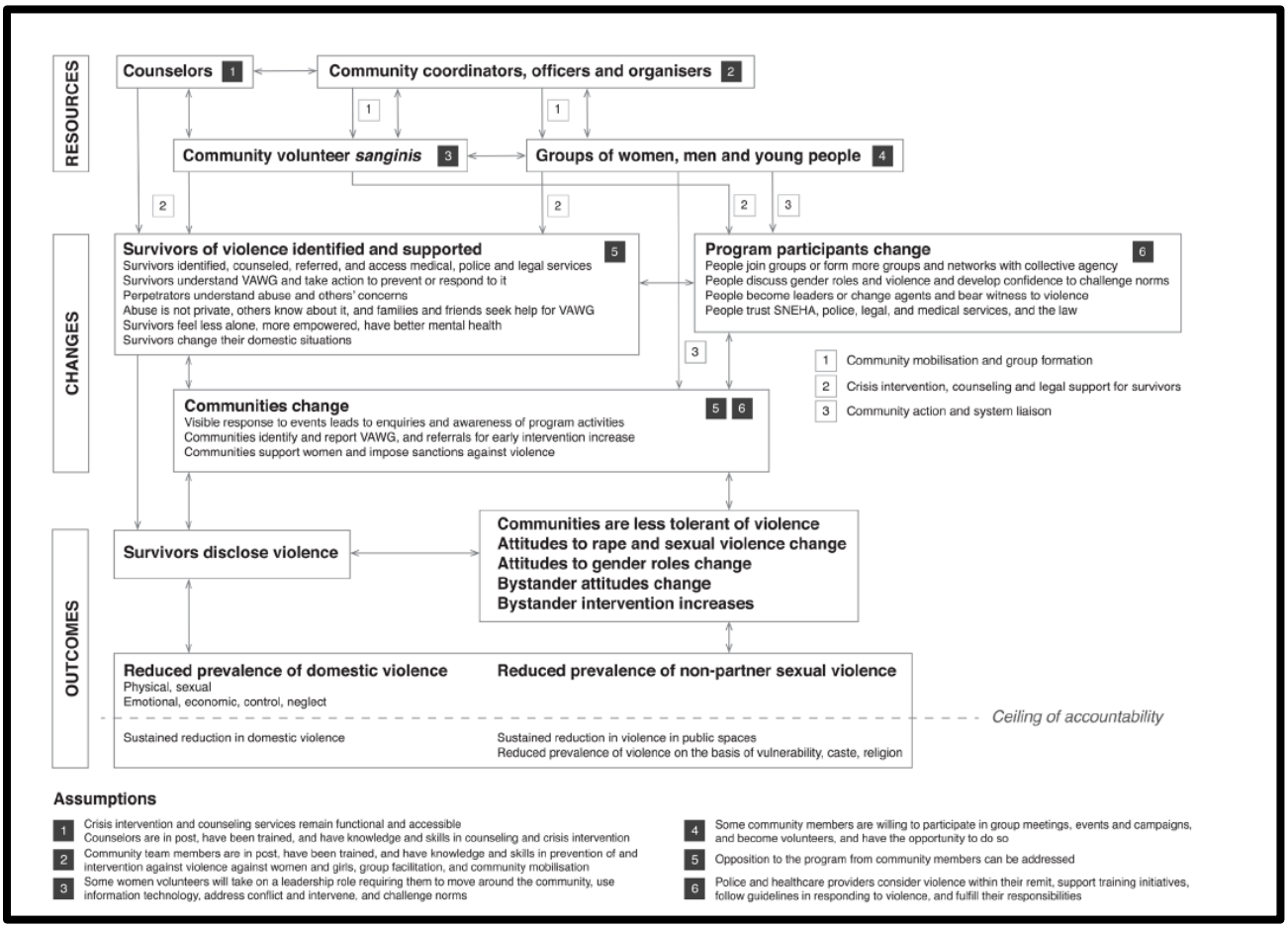
A proposed typology and reporting checklist, Evidence Consortium on Women's Groups, 2020

Classroom, club or collective? Three types of community-based group intervention and why they matter for health, Gram et al., 2020

In this example, Daruwalla et al. (2019) developed a theory of change for an intervention that includes groups as part of a broader community mobilisation strategy to address violence against women (Figure 1). The theory includes a pathway for groups that builds collective agency and challenges social norms, while specifying that groups are open to women, men and young people. The theory of change adheres to recent best practice guidelines (Davies, 2018; De Silva, 2014) by explicitly showing (1) a *ceiling of accountability* separating aspirational outcomes from those outcomes the program is responsible for, and (2) *context and mechanism assumptions* behind each causal link.

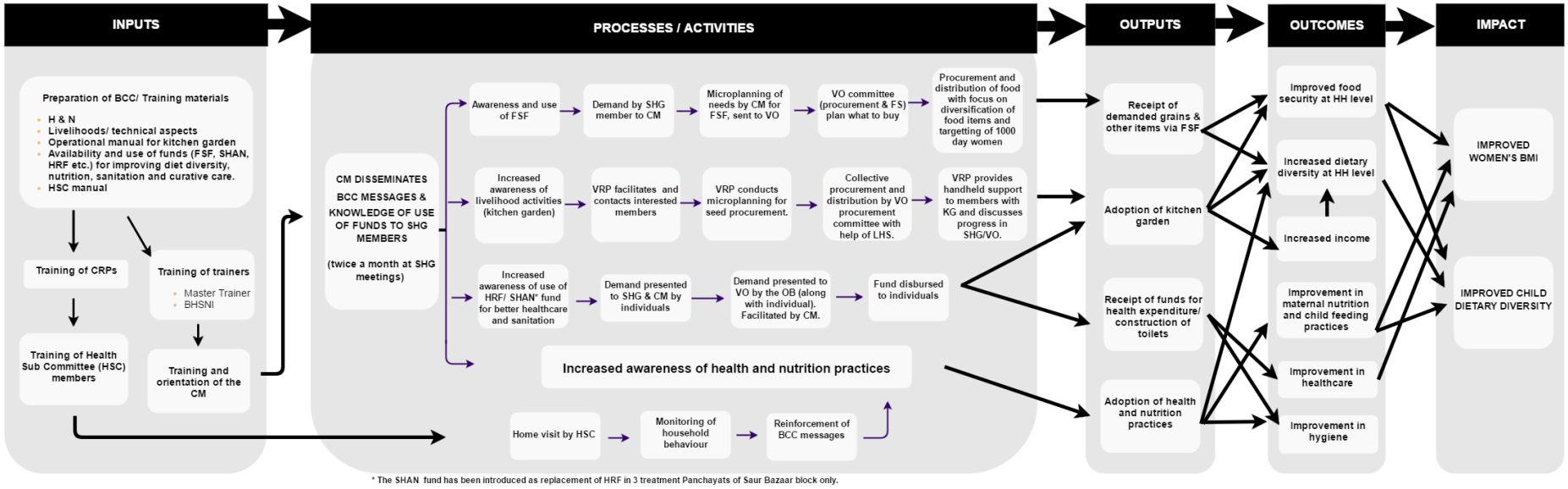
Figure 1. Theory of change for a community mobilisation intervention to address violence against women

(Daruwalla et al., 2019)



In Figure 2 below, Avula et al. (2017) developed a theory of change for an intervention with self-help group members and supply-side convergence that aimed to improve women's anthropometry and child dietary diversity. The theory of change draws from conceptual pathways of how women's groups improve nutrition in South Asia (Kumar et al., 2018) and specifies group-level features (i.e., twice-monthly group meetings; livelihood activities identified through group microplanning) and the health intervention's approach to increasing awareness amongst group members.

Figure 2. Theory of change for self-help groups to improve nutrition outcomes in Bihar, India (Avula et al., 2017)



Identifying health outcome measures

The choice of final health outcomes within a theory of change depends on the intervention design and goals. Health outcomes evaluated in women's group interventions span several domains—most commonly maternal and newborn health, nutrition, sexual health, and mental health (Desai et al., 2020; Orton et al., 2016; Gugerty et al., 2019). The choice of specific health outcome measures within these domains varies widely, however.

We suggest that evaluators consider three questions when identifying health outcomes and measures:

1. *Do the chosen health outcomes have feasible mechanisms of change that align with the intervention theory of change and existing evidence?*

   There is a large body of experimental evidence on women's group interventions and health, with outcome measures as diverse as contraceptive use, mental stress, and health expenditures. Many group-based interventions are premised on mechanisms linked to group membership, such as individual agency, social capital, or collective action. Some taxonomies of behaviour change methods explicitly map approaches that can be used by groups and the wider community to change social support, social norms, environmental conditions, and organisations (Kok, 2016). Linking assumptions regarding mechanisms to existing evidence on groups can sharpen the choice of health outcome. We recommend that evaluators draw from the evidence base on women's groups beyond health, from evidence on mechanisms in groups, and from the public health literature on a specific health outcome.

   For example, an evaluation of the impact of microfinance groups on contraceptive use may hypothesise that improved agency and household-level decision-making will facilitate women's use of oral contraceptive pills (Orton et al., 2016). Evaluations of microfinance groups in a similar setting on their economic and empowerment outcomes can provide insight into the reasonably expected magnitude of effects on women's agency. Recent reviews of evidence of mechanisms can provide insight into the level of confidence in a given mechanism (Gram et al., 2019; Diaz-Martin et al., 2020). Further, evidence from the public health literature on effective interventions to improve contraceptive use can identify enablers and barriers to demand-side interventions, such as financial mechanisms and family-level counselling, that may influence effectiveness of the intervention with groups (Belaid et al., 2016).

Resource Box 3: Identifying mechanisms in group-based interventions

Evidence repository on research on women's groups, Evidence Consortium on Women's Groups

A mixed-methods systematic review of mechanisms, enablers, and barriers, Gram et al., 2019

2. *Are the measures standard or validated?*

   Indicators chosen to measure a health outcome necessarily vary by context, team capacity, and evaluation goals. However, it is important that measures of health outcomes draw on existing tools and guidelines to support consistency. To illustrate, a systematic review of women's groups in India included four evaluations that measured intimate partner violence (Jejeebhoy et al., 2017; Holden et al., 2016; Prillaman, 2017; Yaron et al., 2017). Each study utilised a different outcome and/or index to measure the frequency and type of violence, which limits comparability and quantitative synthesis. Including validated measures of violence against women, such as indicators used in the Demographic Health Surveys or WHO Multi-Country Study on Domestic Violence Against Women, can improve the quality of individual evaluations as well as strengthen comparisons across interventions. We recommend that evaluators include standard

outcome measures and use validated tools specific to the health domain, adapting to the context as necessary.

3. *Are the measures objectively verifiable and/or do they address social desirability bias?*

   Evaluations of interventions that aim to change behaviours typically collect data on a range of self-reported outcomes, such as initiation of breastfeeding or condom use. Evaluators can consider the following options to improve measurement:

   - Inclusion of a "hard" health outcome that is objectively verifiable, such as anthropometry
   - Measures to address social desirability bias in self-reported behavioural outcomes
   - Choosing health outcomes aligned with "benchmark" measures from previously tested, successful interventions in a similar setting

   For example, an evaluation of a nutrition intervention with self-help groups in Bihar included measures of self-reported dietary diversity and anthropometry (Raghunathan et al., 2020). Including anthropometric measures enables stronger comparability of different interventions, such as group-based health programs, cash transfers, and one-to-one counselling by frontline workers. It also provides planners with insights on the theory of change, as improvements in self-reported behaviour do not necessarily translate into improved health outcomes (Acharya et al., 2015; Raghunathan et al., 2020). In addition, particularly for outcomes that necessarily rely on self-reporting, such as violence against women, measures of social desirability bias, accompanying qualitative research, and triangulation with administrative data can improve measurement validity.

2. Measurement scope—group or population?

Community interventions with women's groups typically aim to improve health at a population level. Working with groups may be an efficient way to reach more women, to promote inclusion of the most vulnerable, or to support collective action. Some interventions may focus on group members and their families, while others aim to improve health outcomes amongst the wider community. In a review of 44 evaluations in India, 27 experimental studies reported on population-level outcomes, 15 reported outcomes only among women members or their households, and two studies reported outcomes for separate samples of members and non-members.

Evaluations that report population-level outcomes provide insights to policymakers on intervention coverage and allow comparability to other public health interventions. Reporting on group members alone, without estimates of overall group coverage and sustainability, prevents cost-effectiveness comparisons and interpretation for policy. An exception to this would be if the intervention focuses on improving members' health alone, such as providing group members with access to health insurance or screening exams as part of group activities.

The key difference between measuring outcomes at the population level and measuring them amongst members is the logistics of the sampling approach—not necessarily the sample size. For example, an evaluation of a health intervention with self-help groups to improve maternal and newborn outcomes may choose to sample from group membership lists. However, if the theory of change indicates that a group-based intervention will improve outcomes at a population level because (1) groups have wide population coverage and/or (2) groups diffuse information to other community members or initiate community-wide collective action, sampling from group lists will not allow evaluators to examine these effects. Sampling women through population-based measures, such as household listing or cluster sampling proportionate to population size,

would provide group-level and population-level effects and measures of diffusion. Furthermore, population-based sampling lowers the risk of selection bias, which we discuss in the next section.

In either approach, the evaluation requires a certain number of women who recently had children to detect an effect; the main difference is the underlying sampling frame. If there are only two mothers with young children per self-help group, a larger number of groups will be required than selecting a sample from all eligible women in the given village. In this case, group-level sampling could be more expensive. If groups have low population coverage—and thus are not likely to impact population health—then evaluators may choose to draw a sample from member listing. This approach, however, cannot reliably be used to compare effectiveness with population health interventions or to examine equity. A third option is to sample members and non-members separately, which could allow for estimates of diffusion but not population-level effects. Table 1 provides an overview of advantages and limitations to population- and group-based sampling.

We recommend that evaluators use population-level sampling to allow for reporting of intervention coverage, effects amongst group members, and effects at the population level.

Table 1. Population- versus member-focussed sampling

	Population-based sample	Group members only
Advantages	▪ Comparable effect sizes and cost-effectiveness to other public health interventions ▪ Estimates of actual intervention coverage ▪ Lower risk of selection bias ▪ Can estimate effects on group members and population ▪ Can examine equity	▪ Logistically easier to select respondents through group rosters, if lists are available and updated ▪ Useful where there is low population coverage of groups ▪ Results focus on the direct population of interest (group members)
Limitations	▪ Requires population-level sampling rather than group membership lists ▪ Logistically more challenging if the intervention's focus is group members	▪ Cannot compare effects or costs to other public health interventions ▪ May not have sufficient number of focus population in groups and thus require wider sampling ▪ Group dissolution and drop-outs increase risk of selection bias ▪ Cannot measure spillover effects among the community at large

3. Minimising risk of bias

Quality appraisals in two systematic reviews of the effect of women's groups on health outcomes indicated that at least one third of studies were at high risk of bias (Desai et al., 2020; Orton et al., 2016). The Cochrane Collaboration has developed extensive guidance on addressing risk of bias in evaluations of health outcomes, which we encourage researchers to use as a primary guide (Higgins et al., 2019; Sterne et al., 2016). This section identifies four common sources of bias specific to evaluations of group-based interventions (Table 2).

Baseline confounding is a common risk in non-randomised studies that involve pre-existing groups. For example, implementers may suggest allocation to an intervention area with well-functioning groups that can absorb new interventions, while the control arm has groups of varying strength. However, differences in group characteristics could lead to under- or overestimates of effects. If randomised allocation is not possible, allocation arms should be matched on relevant factors, such as geographic characteristics and outcome of interest (e.g., baseline levels of mortality or undernutrition), as well as group characteristics such as length of functioning, group strength, and drop-out rate. While matching techniques are outside the scope of this

document, we refer readers to the World Bank guide to Impact Evaluation in Practice for detailed recommendations.

Table 2. Mitigating risk of bias in studies of group-based interventions

Source of bias	Group-based interventions	Suggested measures to address bias
Baseline confounding	Allocation is not randomised and allocation arms (i.e., intervention or control arm) differ in strength of groups or baseline levels of outcome	▪ Randomised allocation ▪ Match allocation arms on group and other relevant characteristics (e.g., strength of groups or outcome of interest)
Participant selection	Selection of individuals in surviving groups into the evaluation without accounting for group dissolution	▪ Follow up the same individuals or geographic clusters, whether or not individuals stayed in a group or cluster complied with group intervention ▪ Adjust for selection bias ▪ Sample members and non-members separately ▪ Population-based measurement (e.g., at cluster rather than individual level)
Missing data	Group dissolution and migration leads to missing data on groups and/or participants selected at baseline	▪ Report missing data across arms ▪ Factor attrition into sample size
Selection of reported results	Reporting on a wide range of health outcomes due to lack of formative research, lack of an evidence-based theory of change, or lack of prioritization of a small number of outcomes for a sample size/power calculation	▪ Pre-specify one or a small number of primary outcomes that the evaluation is powered to detect changes in, and produce a registered protocol and analysis plan ▪ Adjust for multiple hypothesis testing when there is more than one primary outcome

Evaluations that select the sample population based on current group membership typically face risks of selection bias at both the group and individual level. Group dissolution is common in large-scale programs, such as India's National Rural Livelihoods Mission (Kochar et al., 2020). If groups dissolve, the sample will only be drawn from better functioning or newer groups. Similarly, if individuals drop out of a group due to poverty, participants will represent a biased sample of women who are better off (Ahmad et al., 2020). Migration may result in missing data and imbalanced attrition across allocation arms. Following the same individuals over time, population-based measurement and attrition analysis are methods for addressing this source of bias. Finally, selection of reported results is a common source of bias in studies without pre-specified primary outcomes. Studies without pre-specified analysis plans may be at risk of selectively reporting or highlighting positive outcomes or changes to health outcome measures during analysis. While not specific to interventions with groups, this source of bias is common in interventions that aim to explore effects on a large number of outcomes. Further, multiple hypothesis testing without appropriate statistical adjustment, while not a source of bias per se, reduces the reliability of results (Sterne et al., 2016).

Resource Box 4: Risk of bias

Cochrane risk-of-bias tool (ROB-2) for randomized trials, Higgins et al., 2019

Cochrane tool for non-randomized studies of interventions, (ROBINS-1), Sterne et al., 2016

Within-study comparisons and risk of bias in international development, Villar & Waddington, 2019

In line with Cochrane's guidance, we recommend that evaluators of group-based interventions publish study protocols with pre-specified primary and secondary outcomes on registries listed in WHO's International Clinical Trials Registry Platform (ICTRP), and append a data analysis plan.

4. Process evaluation and intervention reporting

Evidence syntheses consistently note a lack of reporting on implementation processes and intermediate outcomes, both of which contribute to limited knowledge of the mechanisms that explain effects of interventions with women's groups (Gram et al., 2020; Kumar et al., 2018). Process evaluations can help unpack discrepancies between expected and observed results and identify how the intervention worked, the influence of contextual factors, and the specific mechanisms triggered to achieve outcomes (Craig et al., 2008; White, 2009). A process evaluation can ensure the collection of data along the implementation pathway, allowing researchers both to test a theory of change and identify factors relevant to other interventions. For example, having regular meetings is an assumption inherent in many group-based interventions, yet evidence points to irregular meetings, group dissolution, and limited time for discussion on health as common barriers. Moreover, process evaluations with robust qualitative or mixed-methods components can provide insights into complex social constructs such as community capacity, as well as identify unintended or negative intervention effects (White, 2009; Bonell et al., 2015).

The Medical Research Council guidance for process evaluation of complex interventions suggests measurement and documentation of: implementation structures and processes; mechanisms of impact; and context (Moore et al., 2015). It also provides guidance on the planning, design and conduct, analysis and reporting of process evaluations. Box 2 below provides examples of how process evaluations supported interpretation of impact evaluations as well as raised questions on assumptions and pointed to mechanisms.

Resource Box 5: Process evaluation

Process evaluations of complex interventions: Medical Research Council guidance , Moore et al., 2015

Process evaluations for cluster-randomized trials of complex interventions—proposed framework for design and reporting, Grant et al., 2013

Using participatory process evaluation to understand the dynamics of change in a nutrition education programme, Cornwall, 2014

Readers can consult the LSHTM Centre for Evaluation resource page on process evaluations for additional guidance and examples.

Finally, all evaluations require detailed reporting on interventions to understand impact and transferability. Existing reporting guidelines on intervention processes, such as the TIDieR guidelines, encourage reporting of intervention rationale, content, processes, fidelity, and dose, as well as modification during the course of intervention (Hoffman et al., 2014). Borek et al. (2015), building on a framework for group-based behaviour change developed by Hoddinott and colleagues (2010), propose a detailed checklist specifically for group-based behaviour change interventions (Table 3). Factors such as the total number of sessions and length of sessions are key to understanding implementation intensity of group interventions, as are facilitator characteristics and style.

Box 2. Examples of process evaluations

Explaining the impact of a women's group–led community mobilisation intervention on maternal and newborn health outcomes: the Ekjut trial process evaluation

Rath et al. (2010) conducted a process evaluation of Ekjut's participatory intervention with women's groups, which resulted in large reductions in neonatal mortality in rural Jharkhand and Odisha, India. The process evaluation aimed to (1) describe the intervention's planned content and delivery in practice, and the social context; (2) examine how and why the intervention affected group members and those who did not attend group meetings; and (3) develop and test hypotheses about mechanisms that contributed to the intervention's impact. Evaluators used quantitative and qualitative data sources, including document review, program management data, group interviews, and observations. Data were analysed using a thematic approach, involving both the implementation and evaluation team.

The findings included a description of the intervention context, such as the characteristics of different tribal groups, the geography and terrain, local belief systems, and access to health services. They described the intervention process in detail, including facilitator characteristics, how groups were formed, content and facilitation methods used in the participatory learning and action cycle meetings, and illustrative examples of women's responses to the intervention. In addition to description, the authors identified key enablers through reporting quantitative and qualitative findings, along with illustrative examples. For example, they reported on the proportion of group participants who were men or adolescents, involvement of community health workers, and population coverage of pregnant women and vulnerable groups. Local acceptability and participatory approach to knowledge, skills and developing critical consciousness were explained with qualitative findings and detailed examples. Finally, they reported challenges to the intervention, such as disruptions to group functioning, group dynamics, and health-system barriers to improvements in care-seeking.

The JEEViKA multisectoral convergence pilot in Bihar: A process evaluation report

Avula et al. (2017) conducted a midline process evaluation of a health and nutrition intervention with pre-existing government self-help groups in Bihar. The process evaluation was designed to examine the impact pathways identified in the theory of change (Figure 1 in this document). They identified five broad domains that map to each impact pathway: implementation platforms, training and awareness of roles, implementation processes, exposure of self-help group households to key messages, and exposure to the intervention. They identified a series of research questions corresponding to each domain, such as functionality of platforms, factors that affect delivery, facilitators and barriers to content delivery, intervention coverage, and barriers to adoption of new practices. The authors used quantitative and qualitative data collection, including household surveys, structured interviews, and meeting observations.

The findings reported on the enabling environment for the intervention—particularly the government human resources and institutional structures. They presented detailed findings of implementation processes for both intervention components (behaviour change and system-level convergence). Process findings included data on meeting frequency, women's exposure to the health intervention, and length of discussions on health. Qualitative interviews identified women's lack of interest in maternal nutrition, as members who attended meetings had already completed childbearing. This observation in turn contributed to identifying a mechanism or pre-condition for effective group interventions with self-help groups: choosing issues relevant to group members. Findings on multisectoral convergence indicated poor coordination between frontline health workers and community mobilizers. The midline process evaluation findings pointed to key areas in which to improve implementation—specifically the length of time women spent on health and nutrition discussions and quality of delivery. These findings also helped interpret results of the endline impact evaluation, which indicated improvements in reported childhood dietary diversity but not anthropometry of mothers and children.

The above two points are related to the finding that health and nutrition knowledge among women in treatment areas was not markedly better than women in the comparison areas. And indeed, there did not seem to be an improvement in health and nutrition knowledge over time among women in the treatment areas as compared to those in the control areas. • Given that the health and nutrition BCC is a core component of the intervention, considerable effort will need to be expended to improve its reach and quality in the treatment areas. Without significant differences across treatment and control arms in the knowledge of CMs or of the households, it would be unreasonable to expect differences in household practices or nutritional outcomes

Table 3. Suggested checklist for reporting on group-based behaviour change and health interventions

Adapted from Borek et al., 2015.

Reporting elements	Description
1. Intervention source or development methods	Describes the source (origin) and/or methods used for developing the intervention.
2. General setting	Reports the type of setting where the group sessions were delivered.
3. Venue characteristics	Describes the setup or configuration of the room (or other venue) where the group meetings took place.
4. Total number of group meetings	The total number of group meetings in the program is reported, or it is possible for this to be calculated.
5. Length of group meetings	Reports the length of group meetings (average and/or range).
6. Frequency of group meetings	Reports the frequency of group meetings (i.e., how often they met).
7. Duration of the intervention	Reports the duration of the intervention (i.e., over what period of time group meetings were held).
8. Change mechanisms or theories of change	Describes how the intervention was intended to work by identifying change mechanisms or underpinning theories of behaviour change.
9. Change techniques	Describes the techniques used in group sessions to prompt change. These may be derived from the mechanisms or theories of change, and may use established taxonomies of behaviour change.
10. Meeting content	Describes the content of the meetings in terms of themes or topics covered (i.e., what the sessions were about).
11. Sequencing of meetings	Indicates whether there is a logical (sequential) progression of meeting content or, alternatively, that the content of all meetings is the same (i.e., a repetitive, or "rolling," program, with no particular start or end point).
12. Participants' materials	Reports what materials or tools the participants used during and outside the group meetings.
13. Activities during the meetings	Describes what the participants and the facilitators did during group meetings (i.e., what happened during the meetings—for example, information sharing, discussion, voting). This may be summarized using established taxonomies of behaviour change.
14. Methods for checking fidelity of delivery	Reports methods used to check the fidelity of intervention delivery (i.e., methods used to check if the meetings proceeded as designed).
15. Group composition	Provides information on the composition of the groups in the intervention (i.e., who the participants in the groups were, or whether there were any differences in the participants' characteristics between groups).
16. Methods for group allocation	Describes methods used to allocate participants to different groups.
17. Continuity of participants' group membership	Indicates whether there was continuity in participants' membership in a group throughout the program or if participants could switch between different groups.
18. Group size	Reports the number of participants per group (average and/or range).
19. Number of facilitators	Reports the number of facilitators conducting meetings (i.e., how many facilitators conducted each of the meetings).
20. Continuity of facilitators' group assignment	Indicates whether there was continuity in facilitator's assignment to a group throughout the intervention (i.e., if the same or different facilitators delivered the sessions to each group of participants).
21. Facilitators' professional background	Reports facilitators' professional background, status as a non-professional, or relevant qualifications.

Reporting elements	Description
22. Facilitators' personal characteristics	Reports relevant personal characteristics of the facilitators (i.e., who they were in terms of age, gender, ethnic or cultural background, education level, socioeconomic status).
23. Facilitators' training in intervention delivery	Reports what training in delivering the intervention the facilitators were provided with.
24. Facilitators' training in group facilitation	Reports what training in group facilitation methods the facilitators were provided with (i.e., how to work with and facilitate groups).
25. Facilitators' materials	Reports whether the facilitators were provided with materials and/or written instructions to be used to guide delivery of the meetings.

Conclusion

There is increasing interest in understanding the health effects of interventions with women's groups, as evidenced by the sheer number of evaluations conducted in the past 10 years. Improving the quality and comparability of evaluations will strengthen our understanding of what works, why, where, and for whom. We conclude with a checklist that summarises key recommendations, based on our learning from experimental and quasi-experimental impact evaluations and process evaluations. We welcome continued feedback to add to and adapt this resource document. As interventions are increasingly implemented at scale, utilising evaluation approaches that include population-level data and administrative sources will improve how evidence is generated and used by policymakers.

Box 3: Recommendations to improve evaluations of interventions with women's groups on health outcomes

- Theory of change incorporates group type and health intervention approach with groups
- Choose health outcome measures that are:
  - Validated or standardised
  - Objectively verifiable or address social desirability bias
  - Benchmarked to interventions in a similar setting
- Ensure the sampling frame (group members or population) is aligned with the intervention scope, assumptions, and goals for evidence use
- Address potential sources of bias
  - Register protocol and data analysis plan
- Process evaluation and reporting
  - Mixed-methods process evaluation
  - Track and report group-specific intervention processes

References

Acharya, A., Lalwani, T., Dutta, R., Rajaratnam, J. K., Ruducha, J., Varkey, Wunnava, S., Menezes, L., Taylor, C., & Bernson, J. (2015). Evaluating a large-scale community-based intervention to improve pregnancy and newborn health among the rural poor in India. *American Journal of Public Health, 105*(1), 144–152.

Ahmad, D., Mohanty, I., Irani, L., Mavalankar, D., & Niyonsenga, T. (2020). Participation in microfinance based self help groups in India: Who becomes a member and for how long? *PloS one, 15*(8), e0237519.

Anderson, C. L., Haroon, N., Howlett, M., Kolla, N., Meysonnat, A., Sidhu, A., de Hoop, T., Siwach, G., Desai, S., & Darmstadt, G. (2019). *Women's Empowerment Collectives Portfolio evaluation*. Evans School Policy Analysis & Research Group (EPAR).

Avula, R., Raghunathan, K., Chauhan, T., Chakrabarti, S., Kohli, N., Kumar, N., & Menon, P. (2017). *The JEEViKA multisectoral convergence pilot in Bihar: A process evaluation report*. The World Bank.

Belaid, L., Dumont, A., Chaillet, N., Zertal, A., De Brouwere, V., Hounton, S., & Ridde, V. (2016). Effectiveness of demand generation interventions on use of modern contraceptives in low- and middle-income countries. *Tropical Medicine & International Health, 21*(10), 1240–1254.

Bonell, C., Jamal, F., Melendez-Torres, G., & Cummins, S. (2015). 'Dark logic': Theorising the harmful consequences of public health interventions. *Journal of Epidemiology and Community Health, 69*(1), 95–98.

Borek, A. J., Abraham, J., Smith, R., Greaves, C. J., & Tarrant, M. (2015). A checklist to improve reporting of group-based behaviour-change interventions. *BMC Public Health, 15*(1), 1–11.

Cornwall, A. (2014). Using participatory process evaluation to understand the dynamics of change in a nutrition education programme. *IDS Working Papers, 2014*(437), 1–22.

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2008). Developing and evaluating complex interventions: the new Medical Research Council guidance. *Bmj*, *337*.

Daruwalla, N., Jaswal, S., Fernandes, P., Pinto, P., Hate, K., Ambavkar, G., ... & Osrin, D. (2019). A theory of change for community interventions to prevent domestic violence against women and girls in Mumbai, India. *Wellcome open research*, *4*.

Davies, R. (2018). Representing theories of change: Technical challenges with evaluation consequences. *Journal of Development Effectiveness, 10*(4), 438–461.

De Silva, M. J., Breuer, E., Lee, L., Asher, L., Chowdhary, N., Lund, C., & Patel, V. (2014). Theory of change: A theory-driven approach to enhance the Medical Research Council's framework for complex interventions. *Trials, 15*(1), 1–13.

Desai, S., Misra, M., Das, A., Singh, R. J., Sehgal, M., Gram, L., Kumar, N., & Prost, A. (2020). Community interventions with women's groups to improve women's and children's health in India: a mixed-methods systematic review of effects, enablers and barriers. *BMJ Global Health, 5*(12), e003304.

Dìaz-Martin, L., Gopalan, A., Guarnieri, E., & Jayachandran, S. (2020). Greater than the sum of the parts? evidence on mechanisms operating in women? s groups. Retrieved from: https://faculty.wcas.northwestern.edu/~sjv340/womens_groups.pdf

Evidence Consortium on Women's Groups: Desai, S., de Hoop, T., Anderson, L., Barooah, B., Mulyampiti, L., Obuku, E., Prost, A., & White, H. (2020). *Improving evidence on women's groups: A proposed typology and reporting checklist.*

Gichuru, W., Ojha, S., Smith, S., Smyth, A. R., & Szatkowski, L. (2019). Is microfinance associated with changes in women's well-being and children's nutrition? A systematic review and meta-analysis. *BMJ Open, 9*(1).

Gram, L., Desai, S., & Prost, A. (2020). Classroom, club or collective? Three types of community-based group intervention and why they matter for health. *BMJ Global Health, 5*(12), e003302.

Gram, L., Fitchett, A., Ashraf, A., Daruwalla, N., & Osrin, D. (2019). Promoting women's and children's health through community groups in low-income and middle-income countries: A mixed-methods systematic review of mechanisms, enablers and barriers. *BMJ Global Health, 4*(6), e001972.

Grant, A., Treweek, S., Dreischulte, T., Foy, R., & Guthrie, B. (2013). Process evaluations for cluster-randomised trials of complex interventions: A proposed framework for design and reporting. *Trials, 14*(1), 1–10.

Gugerty, M. K., Biscaye, P., & Leigh Anderson, C. (2019). Delivering development? Evidence on self-help groups as development intermediaries in South Asia and Africa. *Development Policy Review, 37*(1), 129–151.

Higgins, J., Savovic, J., Page, M., & Sterne, J. (2016). Revised Cochrane risk-of-bias tool for randomized trials. *The Cochrane Collaboration*, 2–67.

Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, Altman DG, Barbour V, Macdonald H, Johnston M, Lamb SE, Dixon-Woods M, McCulloch P, Wyatt JC, Chan AW, Michie S. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ*. 2014 Mar 7;348:g1687. doi: 10.1136/bmj.g1687. PMID: 24609605.

Holden, J., Humphreys, M., Husain, S., Khan, S., & Lindsey, S. (2016). *Evaluation of the Madhya Pradesh safe cities initiative 2013–16*. DFID.

Jejeebhoy, S. J., Santhya, K., Acharya, R., Zavier, A., Pandey, N., Singh, S. K., Saxena, K., Rampal, S., Basu, S., & Gogoi, A. (2017). *Empowering women and addressing marital violence through self-help groups: Evidence from rural Bihar.* The Population Council.

Kochar, A., Barooah, B., Jain, C., Closepet, N., Narayanan, R., Sarkar, R., & Shah, R. (2020). *Impact evaluation of the National Rural Livelihoods Programme*. 3ie.

Kok, G., Gottlieb, N. H., Peters, G. J. Y., Mullen, P. D., Parcel, G. S., Ruiter, R. A., ... & Bartholomew, L. K. (2016). A taxonomy of behaviour change methods: an intervention mapping approach. *Health psychology review*, *10*(3), 297-312.

Kumar, N., Scott, S., Menon, P., Kannan, S., Cunningham, K., Tyagi, P., Wable, G., Raghunathan, K. & Quisumbing, A. (2018). Pathways from women's group-based programs to nutrition change in South Asia: A conceptual framework and literature review. *Global Food Security, 17***,** 172–185.

Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, Moore L, O'Cathain A, Tinati T, Wight D, Baird J. Process evaluation of complex interventions: Medical Research Council guidance. *BMJ*. 2015 Mar 19;350:h1258. doi: 10.1136/bmj.h1258. PMID: 25791983; PMCID: PMC4366184.

Orton, L., Pennington, A., Nayak, S., Sowden, A., White, M., & Whitehead, M. (2016). Group-based microfinance for collective empowerment: A systematic review of health impacts. *Bulletin of the World Health Organization 94*(9), 694.

Prillaman, S. A. (2017). *Strength in numbers: How women's groups close India's political gender gap.* Working paper.

Prost, A., Colbourn, T., Seward, N., Azad, K., Coomarasamy, A., Copas, A., Houweling, T. A., Fottrell, E., Kuddus, A., & Lewycka, S. (2013). Women's groups practising participatory learning and action to improve maternal and newborn health in low-resource settings: A systematic review and meta-analysis. *The Lancet, 381*(9879), 1736–1746.

Raghunathan, K., Kumar, N., Gupta, S., Chauhan, T., Kathuria, A. K., & Menon, P. (2020). *Learning together: Experimental evidence on the impact of group-based nutrition interventions in rural Bihar.* International Food Policy Research Institute.

Rath, S., Nair, N., Tripathy, P. K., Barnett, S., Rath, S., Mahapatra, R., Gope, R., Bajpai, A., Sinha, R., & Costello, A. (2010). Explaining the impact of a women's group led community mobilisation intervention on maternal and newborn health outcomes: the Ekjut trial process evaluation. *BMC International Health and Human Rights, 10*(1), 1–13.

Sterne, J. A., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., ... & Higgins, J. P. (2016). ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *bmj*, *355*.

Villar, P. F., & Waddington, H. (2019). Within study comparisons and risk of bias in international development: Systematic review and critical appraisal. *Campbell Systematic Reviews, 15*(1–2), e1027.

Vogel, I. (2012). Review of the use of 'Theory of Change'in international development. *London: DFID*.

Weiss, C. H. (1997). How can theory-based evaluation make greater headway?. *Evaluation review*, *21*(4), 501-524.

White, H. (2009). Theory-based impact evaluation: Principles and practice. *Journal of Development Effectiveness, 1*(3), 271–284.

Yaron, G. (2018). Microfinance for the marginalized: the impact of the Rojiroti approach in India. *Enterprise Development and Microfinance*, *29*(1), 80-93.

The Evidence Consortium on Women's Group (ECWG) is supported by the Bill & Melinda Gates Foundation and aims to address evidence gaps on how groups and collectives can contribute to achieving women's empowerment and well-being as well as understand their implementation models and cost-effectiveness. The consortium is co-led by the American Institutes for Research and Population Council, with partners from the University of Washington, Stanford University, the Campbell Collaboration and Makerere University. To learn more, please visit http://www.womensgroupevidence.org or email info@www.womensgroupevidence.org.