

Carinne Brody
Thomas De Hoop
Martina Vojtkova
Ruby Warnock
Megan Dunbar
Padmini Murthy
Shari L. Dworkin

Economic self-help group programmes for improving women's empowerment

A systematic review

June 2016

Systematic
Review 23

Microfinance



International
Initiative for
Impact Evaluation

About 3ie

The International Initiative for Impact Evaluation (3ie) is an international grant-making NGO promoting evidence-informed development policies and programmes. We are the global leader in funding, producing and synthesising high-quality evidence of what works, for whom, why and at what cost. We believe that better and policy-relevant evidence will make development more effective and improve people's lives.

3ie systematic reviews

3ie systematic reviews appraise and synthesise the available high-quality evidence on the effectiveness of social and economic development interventions in low- and middle-income countries. These reviews follow scientifically recognised review methods, and are peer-reviewed and quality assured according to internationally accepted standards. 3ie is providing leadership in demonstrating rigorous and innovative review methodologies, such as using theory-based approaches suited to inform policy and programming in the dynamic contexts and challenges of low- and middle-income countries.

About this review

Economic self-help group programmes for improving women's empowerment: a systematic review, was submitted in partial fulfilment of the requirements of grant SR4.1049 issued under Systematic Review Window 4. This review is available on the [3ie website](#). 3ie is publishing this report as received from the authors; it has been formatted to 3ie style. This review has also been published by the Campbell Collaboration and is available [here](#).

All content is the sole responsibility of the authors and does not represent the opinions of 3ie, its donors or its board of commissioners. Any errors are also the sole responsibility of the authors. Any comments or queries should be directed to the corresponding author, Carinne Brody, Carinne.brody@tu.edu.

Major funding for this systematic review was provided by UK aid, the Bill & Melinda Gates Foundation and the Hewlett Foundation. A complete listing of all of 3ie's donors can be found on the [3ie website](#).

Suggested citation: 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 Systematic Review 23. London: International Initiative for Impact Evaluation (3ie).

3ie systematic review executive editors: Edoardo Masset and Beryl Leach

Managing editor: Deepthy Menon

Technical Editor: Hugh Waddington

Production manager: Pradeep Singh

Cover design: John F McGill and Akarsh Gupta

Economic self-help group programmes for improving women's empowerment: a systematic review

Carinne Brody
Touro University

Thomas De Hoop
American Institutes for Research

Martina Vojtkova
International Initiative for Impact Evaluation, (3ie)

Ruby Warnock
Touro University

Megan Dunbar
Pangaea Global AIDS Foundation

Padmini Murthy
University of Madras

Shari L Dworkin
University of California, San Francisco

3ie Systematic Review 23

June 2016



Plain language summary

Motivation: Self-help groups (SHGs) are implemented around the world to empower women, supported by many developing country governments and agencies. A relatively large number of studies purport to demonstrate the effectiveness of SHGs. This is the first systematic review of that evidence.

Approach: We conducted a systematic review of the effectiveness of women's economic SHG programs, incorporating evidence from quantitative and qualitative studies. We systematically searched for published and unpublished literature, and applied inclusion criteria based on the study protocol. We critically appraised all included studies and used a combination of statistical meta-analysis and meta-ethnography to synthesize the findings based on a theory of change.

Findings from quantitative synthesis: Our review suggests that economic SHGs have positive effects on various dimensions of women's empowerment, including economic, social, and political empowerment. However, we did not find evidence for positive effects of SHGs on psychological empowerment. Our findings further suggest there are important variations in the impacts of SHGs on empowerment that are associated with program design and contextual characteristics.

Findings from qualitative synthesis: Women's perspectives on factors determining their participation in, and benefits from, SHGs suggest various pathways through which SHGs could achieve the identified positive impacts. Evidence suggested that the positive effects of SHGs on economic, social, and political empowerment run through the channels of familiarity with handling money and independence in financial decision making, solidarity, improved social networks, and respect from the household and other community members. In contrast to the quantitative evidence, the qualitative synthesis suggests that women participating in SHGs perceive themselves to be psychologically empowered. Women also perceive low participation of the poorest of the poor in SHGs due to various barriers, which could potentially limit the benefits the poorest could gain from SHG membership.

Findings from integrated synthesis: Our integration of the quantitative and qualitative evidence suggests there is no evidence for adverse effects of women's SHGs on the likelihood of domestic violence. Women's perspectives in the qualitative research indicate that even if domestic violence occurs in the short term, in the long term the benefits from SHG membership may mitigate the initial adverse consequences of SHGs on domestic violence.

Summary

Background

Women bear an unequal share of the burden of poverty globally due to societal and structural barriers. One way that governments, development agencies, and grassroots women's groups have tried to address these inequalities is through women's SHGs. This review focuses on the impacts of SHGs with a broad range of collective finance, enterprise, and livelihood components on women's political, economic, social, and psychological empowerment.

Objectives

The primary objective of this review was to examine the impact of women's economic SHGs on women's individual-level empowerment in low- and middle-income countries using evidence from rigorous quantitative evaluations. The secondary objective was to examine the perspectives of female participants on their experiences of empowerment as a result of participation in economic SHGs in low- and middle-income countries using evidence from high-quality qualitative evaluations. We conducted an integrated mixed-methods systematic review that examined data generated through both quantitative and qualitative research methods.

Search methods

We searched electronic databases, grey literature, relevant journals and organization websites and performed keyword hand searches and requested recommendation from key personnel. The search was conducted from March 2013–February 2014.

Selection criteria

We included studies conducted from 1980–January 2014 that examined the impact of SHGs on the empowerment of and perspectives of women of all ages in low- and middle-income countries, as defined by the World Bank, who participated in SHGs in which female participants physically came together and received a collective finance and enterprise and/or livelihoods group intervention. To be included in the review, quantitative studies had to measure economic empowerment, political empowerment, psychological empowerment or social empowerment. We also examined adverse outcomes including intimate partner violence, stigma, disappointment, and reduced subjective well-being. We included quantitative studies with experimental designs using random assignment to the intervention and quasi-experimental designs with non-random assignment (such as regression discontinuity designs, “natural experiments,” and studies in which participants self-select into the program). In addition, we included qualitative studies that explored empowerment from the perspectives of women participants in SHGs using in-depth interviews, ethnography/participant observation, and focus groups.

Data collection and analysis

We systematically coded information from the included studies and critically appraised them. We conducted statistical meta-analysis from the data extracted from quantitative experimental and quasi-experimental studies, and used meta-ethnographic methods to synthesize the textual data extracted from the women's quotes in the qualitative studies. We

then integrated the findings from the qualitative synthesis with those from the quantitative studies to develop a framework for assessing how economic SHGs might impact women's empowerment.

Results

We included a total of 23 quantitative and 11 qualitative studies in the final analysis. Initially, we reviewed 3,536 abstracts from electronic database searches and 351 abstracts from the gray literature searches. We found that women's economic SHGs have positive statistically significant effects on various dimensions of women's empowerment, including economic, social and political empowerment ranging from 0.06-0.41 SD. We did not find evidence for statistically significant effects of SHGs on psychological empowerment. We also did not find statistical evidence of adverse effects of women's SHGs. Our integration of the quantitative and qualitative evidence indicates that SHGs do not have adverse consequences for domestic violence. Our synthesis of women's perspectives on factors determining their participation in, and benefits from SHGs suggests various pathways through which SHGs could achieve the identified positive impacts on empowerment. Women's experiences suggested that the positive effects of SHGs on economic, social, and political empowerment run through several channels including: familiarity with handling money and independence in financial decision making; solidarity; improved social networks; and respect from the household and other community members. Our synthesis of the qualitative evidence (key informant interviews and focus groups) also indicates that women perceive there to be low participation of the poorest of the poor in SHGs, as compared to less poor women.

Implications for policy, practice and research

For Policy: SHGs can have positive effects on women's economic, social, and political empowerment. However, we did not find evidence for positive effects on psychological empowerment. These findings indicate that donors can consider funding women's SHGs in order to stimulate women's economic, social, and political empowerment, but the effects of SHGs on psychological empowerment are less clear. Women SHG members perceive that the poorest of the poor participate less than other women. In part, this might be because the poorest of the poor are too financially and/or socially constrained to join SHGs or to benefit from the financial services most often provided through SHGs. Other barriers such as class or caste discrimination might also be present. Poorer or marginalized women may not feel accepted by groups that are made up of wealthier or more well-connected community members. It is important for policy makers to identify ways to build in support and reduce barriers for individual women who want to participate in SHGs but who do not have the financial resources or freedoms to join.

For Practice: We do not find evidence for adverse effects of women SHGs on domestic violence based on the integration of the quantitative and the qualitative evidence. Although there may be adverse consequences in the short term, analysis of women's reports suggest that SHGs do not contribute to increases in domestic violence in the long term. Furthermore, participation of the poorest of the poor in SHGs may be stimulated by incentives. These incentives could be financial, for example, by giving the poorest of the poor the opportunity to participate without a savings requirements, or non-financial, for example, by stimulating the husbands or mothers-in-law of the poorest of the poor to let their spouses and daughters-in-law participate in SHGs or conducting outreach activities to marginalized groups. As new programs are implemented in different contexts, it is also important that

program designs are tailored to the local settings in ways that allow them to evolve over time. This review has shown that one-size does not fit all, and while it is important to take best practices across programs for implementation, this means that flexibility is required to adapt programs successfully for the greatest impact in women's lives.

For Research: There is a need for more rigorous quantitative studies that can correct for selection bias, spillovers and the difficulties of measuring empowerment. There is also a need for more research, focused on examining possible factors that mediate and/or moderate the impact of SHGs on women's empowerment to further understand the pathways or mechanisms through which SHGs impact empowerment. For the latter it is crucial to conduct rigorous qualitative research in addition to rigorous quantitative research. Whereas quantitative research is useful in understanding certain aspects of the impact of SHGs on empowerment, qualitative studies could show us more nuanced ideas about how to measure empowerment. Importantly, both quantitative and qualitative studies need to describe more fully the various components of the SHGs being studied. Greater detail in the description of the program design will help in determining moderating factors in the design of SHGs.

Contents

Plain language summary	I
Summary	II
List of tables and figures	VII
1. Background	1
1.1. Description of the problem	1
1.2. Description of the intervention	1
1.3. How the intervention might work	2
1.4. Why it is important to do this review	4
2. Objectives of the review	8
3. Methods	9
3.1. Criteria for including studies in the review	9
3.2. Search methods for identification of studies	12
3.3. Data collection and analysis	15
3.4. Deviations from the protocol	25
4. Results	27
4.1. Results of the search	27
4.2. Description of included studies	30
4.3. Critical appraisal of included studies	42
4.4. Synthesis of quantitative studies	46
4.5. Synthesis of qualitative studies	70
4.6. Integrated synthesis	81
5. Discussion	85
5.1. Summary of main results	85
5.2. Overall completeness and applicability of evidence	85
5.3. Quality of the evidence	87
5.4. Limitations and potential biases in the review process	87
5.5. Agreements and disagreements with other studies or reviews	90
6. Authors' conclusions	91
6.1. Implications for practice and policy	91
6.2. Implications for research	92
7. Acknowledgments	94
References	95
Appendix 1: Data extraction form	109
Appendix 2: Full search strategy	112
Appendix 3: Search diary	117
Appendix 4: Reasons for exclusion of marginal studies	121
Appendix 5: Qualitative study quality assessment	124
Appendix 6: Quantitative risk of bias assessment tool	127
Appendix 7: Overview of risk of bias assessment of included quantitative studies ..	133
Appendix 8: Detailed risk of bias assessment of included quantitative studies	135
Appendix 9: Quality assessment for included qualitative studies	143

Appendix 10: Procedures for calculating effect sizes.....	145
Appendix 11: Additional forest plots.....	148
Appendix 12: Additional quotes by theme	153
8. Contribution of authors.....	159
9. Declarations of interest	160
10. Sources of support.....	161

List of tables and figures

Table 4. 1: Reasons for exclusion of studies	29
Table 4. 2: Sources and publication types for included studies	29
Table 4. 3: Summary of included quantitative studies.....	32
Table 4. 4: Summary data on qualitative studies	39
Table 4. 5: Measurement of women's economic empowerment	46
Table 4. 6: Summary of effects of SHGs on economic empowerment.....	53
Table 4. 7: Measurement of women's social empowerment	54
Table 4. 8: Summary effects of SHGs on women's mobility	61
Table 4. 9: Summary effects of SHGs on women's family-size decision-making power.....	62
Table 4. 10: Measurement of political empowerment	63
Table 4. 11: Measurement of psychological empowerment.....	65
Table 4. 12: Measurement of intimate partner violence	66
Table 4. 13: Summary effects of SHGs on women's political and psychological empowerment and intimate partner violence.....	68
Table 4. 14: Summary of evidence in qualitative studies	71
Table 4. 15: Summary of findings from qualitative studies.....	72
Figure 1. 1: Economic self-help groups and empowerment causal pathway.....	3
Figure 4. 1: Study search	28
Figure 4. 2: Risk of bias assessment of quantitative studies.....	43
Figure 4. 3: Summary of quality appraisal of qualitative studies	44
Figure 4. 4: The effects of women's self-help groups on women's economic empowerment (randomized controlled trials).....	49
Figure 4. 5: The effects of women's self-help groups on women's economic empowerment (quasi-experimental evaluations)	50
Figure 4. 6: Effects of women's self-help groups on women's economic empowerment (RCTs and quasi-experimental evaluations with a medium risk of selection- bias).....	52
Figure 4. 7: The effect of women's self-help groups on women's social empowerment (randomized controlled trials).....	56
Figure 4. 8: Effects of women's self-help groups on women's social empowerment (quasi- experimental evaluations)	57
Figure 4. 9: Effects of women's self-help groups on women's mobility (quasi-experimental evaluations)	58
Figure 4. 10: Effects of women's self-help groups on women's social empowerment (RCTs and quasi-experimental evaluations with a medium risk of selection-bias)	59
Figure 4. 11: Effects of women's self-help groups on women's mobility	60
Figure 4. 12: Effects of women's self-help groups on women's family-size decision-making	60
Figure 4. 13: Effects of women's self-help groups on political empowerment	64

Figure 4. 14: Effects of women's self-help groups on psychological empowerment.....	65
Figure 4. 15: Effects of women's self-help groups on intimate partner violence.....	67
Figure 4. 16: Funnel plot of economic empowerment outcome	69
Figure 4. 17: Funnel plot of social empowerment outcome	69
Figure 4. 18: Revised theory of change.....	84
Figure 10. 1: Estimation of odds-ratios.....	146
Figure 11. 1: Meta-Analysis for determining the effect of women's self-help groups on women's economic empowerment based on quasi-experimental evaluations with a high risk of selection-bias	148
Figure 11. 2: Meta-Analysis for determining the effect of women's self-help groups on women's economic empowerment based on quasi-experimental evaluations with a medium risk of selection-bias.....	148
Figure 11. 3: Meta-Analysis for determining the effect of women's self-help groups on women's economic empowerment based on RCTs and quasi-experimental evaluations with a medium risk of selection-bias that focus on SHGs with a training component	149
Figure 11. 4: Meta-Analysis for determining the effect of women's self-help groups on women's economic empowerment based on RCTs and quasi-experimental evaluations with a medium risk of selection-bias that focus on SHGs without a training component	149
Figure 11. 5: Meta-Analysis for determining the effect of women's self-help groups on women's social empowerment based on quasi-experimental evaluations with a high risk of selection-bias.....	150
Figure 11. 6: Meta-Analysis for determining the effect of women's self-help groups on women's social empowerment based on quasi-experimental evaluations with a medium risk of selection-bias.....	150
Figure 11. 7: Meta-Analysis for determining the effect of women's self-help groups on women's family-size decision making power based on quasi-experimental evaluations with a high risk of selection-bias.....	151
Figure 11. 8: Meta-Analysis for determining the effect of women's self-help groups on women's family-size decision-making based on RCTs and quasi-experimental evaluations with a medium risk of selection-bias that focus on SHGs with a training component	151
Figure 11. 9: Meta-Analysis for determining the effect of women's self-help groups on women's mobility based on RCTs and quasi-experimental evaluations with a medium risk of selection-bias that focus on SHGs with a training component	152

1. Background

1.1. Description of the problem

Women bear an unequal share of the burden of poverty globally due to societal and structural barriers. According to economist and Nobel laureate Amartya Sen (2001), women worldwide have less access to “substantive freedoms” such as education, employment, health care, and democratic freedoms. First, girls are enrolled in school at lower rates than boys, resulting in women making up more than two-thirds of the world's illiterate adults (UNESCO, 2013). Second, women experience unequal access to health care starting from birth and throughout their reproductive years (WHO, 2007). Third, women are missing from all levels of government—local, regional, and national (Lopez-Claros, 2005). Women also have fewer economic freedoms. In sub-Saharan Africa, only 16 to 18 per cent of loans issued to small and medium businesses are issued to women owners; and in South Asia, only 6 per cent (IFC, 2014). In addition, in many countries, women cannot own land. In South and Southeast Asia, women comprise more than 60 per cent of the agricultural labor force. However, in India, Nepal, and Thailand less than 10 per cent of women farmers own land (FAO, 2008). These facts describe what economists call the *feminization of poverty*. This phrase is meant to capture women's unequal share of poverty, in terms of both wealth and choices and opportunities (Sen, 2001).

One way that governments, development agencies, and grassroots women's groups have tried to address these inequalities is through women's economic self-help groups (SHGs). The basic assumptions undergirding these income-generating group programs are that giving women access to working capital can increase their ability to “generate choices and exercise bargaining power as well as develop a sense of self-worth, a belief in one's ability to secure desired changes, and the right to control one's life” (UN, 2000). SHGs of women could facilitate these goals through the development of social capital and mobilization of women (IFAD, 2003).

1.2. Description of the intervention

SHGs, also known as mutual aid or support groups, are small voluntary groups that are formed by people related by an affinity for a specific purpose who provide support for each other. They are created with the underlying assumption that when individuals join together to take action toward overcoming obstacles and attaining social change, the result can be individual, and/or collective empowerment. SHG members typically use strategies such as savings, credit, or social involvement as instruments of empowerment. The types of SHGs that exist in developing countries are numerous and can include economic, legal, health, and cultural objectives.

The canonical economic SHG model starts with an initial period of collective savings in the name of the group to facilitate intragroup lending. The basic idea underlying this model is that groups then gradually take larger loans, for example, from banks. In addition, SHGs often provide support in the form of training, which can take multiple forms. Trainings can, for example, focus on entrepreneurial skills, women's rights, political participation, basic education, and justice (Van Kempen, 2009). SHGs can be linked directly with banks or can function through non-governmental organizations (NGOs) and tend to be more fundamentally grassroots in nature than the many microfinance institutes (MFIs) that now exist worldwide. Although SHGs share some important characteristics, there are major

differences across SHGs as well. For example, Thorp et al. (2005) suggest that some SHGs focus on resolving market failures, such as saving and credit constraints, while others put a stronger emphasis on rights, for example group members' rights to access resources or political participation.

India and other countries in South and Southeast Asia have a long history of SHG activity. South Asia's largest and perhaps most well-known program is the Self-Help Group-Bank Linkage Program (SBLP). This Indian program was started in 1992 and has rapidly expanded since then. In 2009, the SBLP covered approximately 86 million poor households in 6.1 million saving-linked SHGs and 4.2 million credit-linked SHGs. The SBLP is best known for its expansive outreach and high repayment rates of over 95 per cent. The literature suggests that the program has been effective at targeting poor women and is associated with improvements in household income, livestock ownership, savings and households' ability to withstand economic shocks (Sinha, 2008). In addition, the program might have contributed to improvements in women's decision-making power, control over household resources, and participation in the public sphere (Sinha, 2008). In other parts of the world, such as sub-Saharan Africa and Latin America, the South Asian model has been adapted to match the cultural and social context in those specific settings. For example, SHGs in sub-Saharan Africa, such as Jeunes sans Frontières in Burkina Faso, have a stronger emphasis on HIV/AIDS than SHGs in Asia. African SHGs may thus have contributed to overcoming the stigma surrounding HIV/AIDS in sub-Saharan Africa (Nguyen, 2005).

The majority of SHGs target women with the explicit goal of empowering them. For example, the SHG model "was introduced as a core strategy to achieve empowerment in the Ninth Plan (1997-2002) with the objective to 'organize women into Self-help group [sic] and thus mark the beginning of a major process of empowering women' (Planning Commission, 1997). Jakimow and Kilby (2006) argue, however, that in practice the South Asian SHG model is often focused on solving market failures, by emphasizing credit and saving, rather than empowering women.

This review focuses on SHGs that offer women a collective finance, enterprise, and/or livelihood component. Collective finance and enterprise can include savings and loans, group credit, collective income-generation, and micro-insurance. Livelihood interventions can include life skills training, business training, financial education, and labor and trade group organizing.

1.3. How the intervention might work

Many different perspectives, definitions, measures, and outcomes have been associated with women's empowerment. The growing literature presents different definitions of empowerment, and no one definition seems to be universally accepted. For example, women's empowerment is used interchangeably with other terms such as women's autonomy, status, and agency. These terms have subsequently been measured in different ways. For example, women's autonomy has been measured by assessing the degree to which women participate in decision-making in their households (Upadhyay, 2005) or by determining women's mobility (Malhotra, 2002). Additional challenges in defining and measuring women's empowerment include variations in the cultural context that affect how empowerment may occur. For example, women's mobility may be a central issue to women's empowerment in one setting and a peripheral issue in another. Differences in the

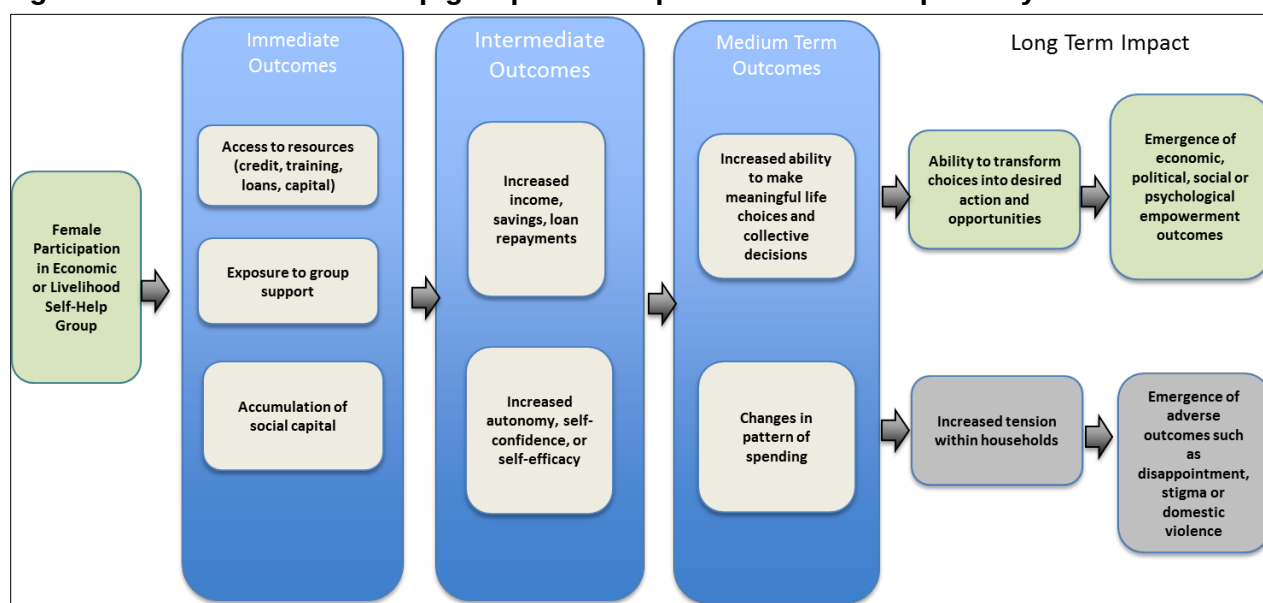
approach to measure empowerment and contextual differences complicate the process of defining whether different measures of empowerment can be considered part of the same construct in this systematic review. We will discuss this issue in detail in later stages of this review.

Nonetheless, much of the research agrees that empowerment is a process and an outcome that can occur at multiple levels and within different dimensions. After the International Conference on Population and Development (United Nations Population Information Network & United Nations Population Fund, 1996), the UN delineated five major components of empowerment:

1. Women's sense of self-worth
2. Women's right to have and to determine choices
3. Women's right to have access to opportunities and resources
4. Women's right to have the power to control their own lives, both within and outside the home
5. Women's ability to influence the direction of social change to create a more just social and economic order, nationally and internationally.

One of the more comprehensive and broadly cited definitions of empowerment comes from a study by Kabeer (1999, p. 437) who states that empowerment is “the expansion in people’s ability to make strategic life choices in a context where this ability was previously denied to them; a process that entails thinking outside the system and challenging the status quo, where people can make choices from the vantage point of real alternatives without punishingly high costs.” This definition is reflected in our theory of change underlying economic SHGs, which includes resources (for example, increased income, savings, and loan repayments), agency (for example, increased autonomy, self-confidence, or self-efficacy), and achievements (for example, ability to transform choices into desired action and opportunities) (Kabeer, 1999). We based our review on the theory of change underlying economic SHGs as depicted in Figure 1.1.

Figure 1. 1: Economic self-help groups and empowerment causal pathway



Source: authors.

Based on the literature, we hypothesized that women's participation in economic and livelihood SHGs would enable women to gain access to resources in the form of credit, training, loans, or capital. As a result, women SHG members might experience an increase in income, savings, and/or loan repayments. In addition, participants would be exposed to group support. As a result of group support, women SHG members might experience increased feelings of autonomy, self-confidence, and self-efficacy. Following increased financial stability and self-confidence, women SHG members might then be able to make meaningful life choices, and their patterns of spending and savings might change. As a result of these changes, women SHG members might experience an increased ability to transform their choices into desired actions, which would lead to the emergence of economic, political, social, and psychological empowerment (Eyben, Kabeer & Cornwall, 2008). The potential for these changes to occur are dependent upon "context, commitment and capacity" (Kabeer, 2005).

Empowerment studies have lent credence to the concept that women can and should be central actors in social and economic development, but empowerment of an individual or a small group alone might invoke negative reactions when familial, community, and structural factors have not yet adjusted to women's changing roles. Intimate partner violence, for example, has been shown to increase when women's economic empowerment is not complemented with additional interventions that focus on mitigating the potential adverse consequences at the household and community level (Ahmed, 2005; Dalal, 2011). Thus, several studies recommend complementing interventions with an emphasis on empowering women with interventions that focus on changing the gender norms of men (for example, Barker & Schulte, 2010; Dworkin et al., 2011; Dworkin, Forthcoming).

Studies also suggest that increasing women's monetary contributions to the family without also taking into account the upheaval this might cause with respect to expected gender and domestic responsibilities can lead to increased household and community tensions and decreased emotional well-being for women (Ahmed 2005; Ahmed & Chowdhury, 2001; De Hoop et al., 2014). Short- and long-term backlash tendencies are, therefore, important to consider when examining the impacts of SHGs on empowerment.

Numerous factors can modify the pathways described here. For example, the literature highlights that empowerment can occur at the individual and collective levels (Eyben, Kabeer & Cornwall, 2008). *Individual empowerment* refers to changes that occur within an individual. *Collective empowerment* refers to structural changes at the societal level in terms of how relationships and institutions impact households and individuals. Although SHG participation might lead to improved self-efficacy of an individual (individual empowerment), the systematic marginalization of the group might remain unchanged (collective empowerment). Hence, individual empowerment does not necessarily result in collective empowerment. The economic climate, program fidelity, role of the facilitator, and underlying race, ethnicity, class and/or caste issues can also affect how program benefits are realized.

1.4. Why it is important to do this review

Today, women's empowerment is considered an essential component of international development and poverty reduction. The concept of women's empowerment has gained increased attention over the past two decades. This concept first held international prominence at the International Conference on Population and Development in Cairo in

1994 and then again at the Fourth World Conference on Women, Beijing, 1995. But the central role of women in development originated during grassroots movements that commenced years earlier.

The international conferences at Cairo and Beijing announced the shift from thinking of women as targets for fertility control policies to acknowledging women as autonomous agents with rights. As a result of these conferences, a broad assessment of women's empowerment throughout the United Nations (UN) system was undertaken. By 2000, when 189 UN member states created eight poverty reduction targets called the Millennium Development Goals, they agreed that "promoting gender equality and empowering women" deserved to be included as a stand-alone goal in addition to the other health and education-related targets (UNDP, 2010). In addition, the UN now assesses the different implications of development planning for women and men and integrates poverty eradication strategies into programs for women (African Women's Development and Communication Network, 2010).

The international conferences at Cairo and Beijing helped shift resources and ideologies toward women's role in development, but the emergence of women's empowerment as a central concept in development was the result of earlier grassroots movements aimed at empowering disenfranchised communities with women playing a central role. Grassroots organizing included the formation of SHGs, which became a central ground for women's activism and participation and helped to shape the changing development landscape in South Asia. Nowadays SHGs are among the most popular programs that aim to stimulate the empowerment of women in South Asia (Jakimow & Kilby, 2006). Although SHGs have a less prominent history in low-and middle-income countries outside South Asia, the formation of SHGs has also diffused to countries in other parts of Asia, Sub-Saharan Africa, and Latin America.

The concept of the SHG as a catalyst for change in developing countries was based on the self-help approach pioneered in India in the early 1980s. It emphasized high levels of group ownership, control, and management concerning goals, processes, and outcomes. It has been argued that the very process of making decisions within the group is an empowering process and can lead to broader development outcomes such as the greater participation of women in local governance and community structures (Mayoux, 1998). For example, in case studies of women's cooperatives in rural Nigeria and rural India, women who were engaged in cooperative activities appeared to be more productive and had higher levels of economic well-being, than non-members (Amaza, Kwagbe & Amos, 1999; Datta & Gailey, 2012).

As these smaller SHGs became successful, larger umbrella organizations emerged with the goal of harnessing the energy of smaller groups and advocating for the rights of the poor and of women on the global stage. One example of an umbrella organization is the Self Employment Women's Association (SEWA), which was launched in the state of Gujarat, India, by female garment workers, who first met in a park to discuss their working conditions and eventually organized into a trade union. This project, which was launched in 1972, has included thousands of women and their families (Narayan et al., 2000).

Following the global recognition of the critical role of women in poverty reduction strategies, a wave of microfinance programs and other livelihood support interventions were implemented worldwide, specifically targeting rural women and women's SHGs. As

discussed above, a large majority of these programs focus explicitly on empowerment, although the emphasis is sometimes on resolving market failures.

We based our review on the understanding that a great deal of evidence about women's SHGs has already been generated from quantitative and qualitative research, much of which might be useful in informing policy and practice.

Several systematic reviews focus on the impact of microfinance on economic well-being. First, Duvendack and colleagues (2011) reviewed the evidence of the impact of microfinance on the well-being of poor people. The authors found only limited evidence that microfinance improves economic well-being, but felt limited by the lack of rigorous impact evaluations on microfinance. Second, a systematic review by Stewart and colleagues (2010) on the impact of microfinance on poor people in Sub-Saharan Africa came to similar conclusions with respect to microcredit. The authors concluded, however, that based on the evidence they included in their review, micro-savings appeared to be more effective in improving the well-being of poor people. Following this conclusion, the authors called for more rigorous evidence on the impact of microsavings programs. Third, Stewart and colleagues (2011) reviewed whether microcredit, microsavings, and microleasing serve as effective financial inclusion interventions enabling poor people, and especially women, to engage in meaningful economic opportunities in low- and middle-income countries. The authors found mixed results once again. In some cases, microcredit and microsavings reduced poverty but not in all circumstances or for all clients. The authors also showed that there was not enough evidence to say that microfinance interventions targeting women exclusively were more successful at reducing poverty than those targeting both men and women.

The findings of these reviews stand in stark contrast to the prevailing positive view about the impact of microfinance on poverty reduction before these reviews and a number of randomized controlled trials (RCTs) were conducted. The prevailing positive view was mostly based on anecdotal evidence and studies that were vulnerable to selection bias (Roodman, 2011). Both donor and nongovernmental organizations promoted microfinance on the basis of an understanding that it reduced poverty and empowered women (White & Waddington, 2012). However, new rigorous evidence from RCTs on the impacts of microcredit on poverty reduction and women's empowerment suggests that the effectiveness of microcredit is at best modest (Attanasio et al., 2015; Augsburg et al., 2015; Banerjee et al., 2015; Crepon et al., 2015).

A recent systematic review on the impact of microcredit on women's bargaining power also suggests that the prevailing positive view on the effects of microcredit on women's empowerment might be overstated (Vaessen et al., 2014). The evidence from the most rigorous studies in that review, including those based on RCTs and credible quasi-experiments, suggested there was no evidence for a causal link between microcredit and women's control over household spending.

There are, however, several mechanisms through which SHGs can improve women's empowerment. Apart from the economic channel, it is also important to focus on the potential effects that group-support and training might have on women's empowerment. We focused on both of these mechanisms in the theory of change described above.

The reviews cited previously were restricted to microcredit and microsavings interventions and did not comprehensively review and synthesize the evidence on the impact of SHGs that included collective finance, enterprise, and/or livelihoods components. In addition, the reviews did not comprehensively cover a range of key empowerment outcomes such as decision making within households, feelings of self-confidence or autonomy, or the ability to exercise control over family planning. Although Vaessen et al.'s review is the only one with an explicit focus on women's empowerment, the review does not focus exclusively on SHGs, covers only microcredit interventions, and does not synthesize empowerment outcomes other than women's control over household resources.

The current review focuses on quantitative studies evaluating the impact of SHGs with a broad range of collective finance, enterprise, and livelihood components on political, economic, social, and psychological empowerment in addition to women's control over household resources. This systematic review thus goes beyond determining the effects of microcredit on women's empowerment to ensure that we learn about the credit, saving, group support, and training components of women's SHGs.

In order to identify some of the pathways and moderators, we also included qualitative studies of women's perceptions of the barriers and facilitators to women's empowerment within SHGs. We recognize that heterogeneity in the design and implementation of SHGs makes it difficult to interpret the existing evidence on the impact of SHGs on women's empowerment. Our systematic review assesses the effects of women's SHGs and the pathways and moderators to explain these effects by using a mixed-methods evidence synthesis as in the systematic review on the effects of farmer field schools (Waddington et al., 2014).

The protocol of this study is available through the Campbell Collaboration Library of Systematic Reviews (Brody et al., 2014).

2. Objectives of the review

The primary objective of this review is to examine the impact of women's economic SHGs on individual-level empowerment for women in low- and middle-income countries, using evidence from rigorous quantitative impact evaluations (review objective 1).

The secondary objective of this review is to examine the perspectives of female participants on factors determining their participation in, and benefits from, economic SHGs in low- and middle-income countries using evidence from high-quality qualitative evaluations (review objective 2).

Finally, this review aims to refine the theory of change introduced in section 1 that describes how women's economically oriented SHGs lead to women's empowerment using evidence drawn from both rigorous quantitative impact evaluation studies and qualitative studies about perspectives of women who are SHGs participants.

3. Methods

3.1. Criteria for including studies in the review

We conducted an integrated mixed-methods review that examines data generated through both quantitative and qualitative research methods. We believe this study design will enhance the review's utility and impact for practitioners and policymakers. This approach allowed us to capture a broader range of evidence than a review of quantitative studies alone so that we could answer relevant policy questions more comprehensively.

We included studies in the review that fulfilled the following criteria.

3.1.1. *Participants*

SHG participants included women of all ages in low- and middle-income countries, as defined by the World Bank categorization of low- and middle-income countries, at the time the data were collected. Women's SHGs and SHGs in which participation was either limited exclusively to women or, if this was not the case, in which impacts on women were assessed separately from men, were included. In contrast, studies were excluded in which impacts were not disaggregated by gender and/or self-help groups were comprised exclusively of men.

3.1.2. *Interventions: type of women's self-help group programs*

We included studies on SHGs in which female participants physically came together and received a collective finance and enterprise and/or livelihoods group intervention:

- We defined SHGs, also known as mutual aid or support groups, as those groups that involved people who provide support for each other and/or are created with the underlying assumption that when individuals join together to take action toward overcoming obstacles and attaining social change, individual, and/or collective empowerment can result.
- We planned to examine those groups that were initiated by an external agency (that is, a development organization or research group) as well as those that had come into existence without any direct external involvement. In practice, however, all included studies focused on groups that were initiated by an external agency.
- SHGs needed to receive an economic intervention that included or contained the following components: collective finance and enterprise¹ (such as savings and loans, group credit, collective income-generation, micro-insurance) and/or livelihoods

¹ An example of a collective finance intervention is SaveAct in South Africa, which allows members of the community to voluntarily form a group and save money in the form of share purchases. The group also contributes monthly to a Social Fund to assist members in times of emergency or family crisis, such as a death in a member's family (SaveAct.org, 2013).

interventions (such as life skills, capacity-building, business training, financial education, labor and trade group organizing).²

- We excluded studies evaluating individual self-help or group programs that were not explicitly designed as self-help programs or did not have a collective finance, enterprise, or livelihoods intervention component.

3.1.3. Outcomes

Primary outcomes

To be included in the review, studies had to measure at least one of the following empowerment outcomes.³

Economic empowerment: We defined women's economic empowerment as the ability of women to access, own, and control resources. It could be measured in a variety of ways, using outcome indicators such as income generation by women, female ownership of assets and land, expenditure patterns, degree of women participation in paid employment, division of domestic labor across men and women, and control over financial decision making by women.

Political empowerment: We defined political empowerment as the ability to participate in decision making focused on access to resources, rights, and entitlements within communities. It could be measured using indicators such as awareness of rights or laws, political participation such as voting, the ability to own land legally, the ability to inherit property legally, and the ability to gain leadership positions in the government.

Social empowerment: We defined social empowerment as the ability to exert control over decision making within the household. Measures included women's mobility or freedom of movement, freedom from violence, negotiations and discussion around sex, women's control over choosing a spouse, women's control over age at marriage, women's control over family size decision making, and women's access to education.

Psychological empowerment: We defined psychological empowerment as the ability to make choices and act on them. It could be measured using outcome indicators such as self-efficacy or agency; feelings of autonomy; and sense of self-worth, self-confidence, or self-esteem.

The definition of the outcome measures shows that empowerment is a broad concept even when we divide it into four empowerment constructs. Furthermore, study authors of primary studies use a large number of different operational definitions to measure economic, social, political, and psychological empowerment. The large number of outcome measures to operationalize empowerment is not surprising, since the concept is difficult to define. Nonetheless, we had to be careful in grouping outcome variables when we were not certain whether these outcome variables measured the same construct. At the same time, the

² An example of an individual livelihoods intervention was the *Neang Kongrey Stoves* project in Cambodia, which offered training program to three groups of local potter women on how to produce improved cook stoves (World Bank, 2009).

³ Sources: Malhotra, Schuler & Boender, 2002; Mayoux, 1998; Eyben, Kabeer & Cornwall, 2008.

literature on measuring empowerment suggests that empowerment should be considered a latent construct that cannot be measured using one specific outcome variable. Thus, several researchers use an index to measure empowerment (Pitt, Khandker & Cartwright, 2006; Bali Swain & Wallentin, 2009). These indices suggest that different operational definitions to measure empowerment can be considered part of the same construct. For example, several studies construct indices based on variables that measure different elements of women's bargaining power, mobility, family-size decision-making, and political, as well as psychological empowerment (Pitt et al., 2006; Bali Swain & Wallentin, 2009). Nonetheless, we took seriously the concern that different operational definitions of empowerment cannot always be considered part of the same construct. Thus, we used an iterative approach in the definition of our outcome measures. First, we grouped outcome variables under economic, social, psychological, and political empowerment. Second, we synthesized the evidence on the effects of women's SHGs on these four constructs of women's empowerment under the assumption that it is appropriate to group the outcome variables under the same construct. Third, we analyzed the robustness of the results to excluding studies with outcome measures that might not measure the same construct as the other outcome variables.

Secondary outcomes

We also examined spillover effects from women's SHG participants to nonparticipating women in the same communities on the same outcomes.

In addition, we examined adverse outcomes including:

- Intimate partner violence.
- Stigma.
- Disappointment.
- Reduced subjective well-being.

3.1.4. Study types

To answer our review questions, we included studies with study designs and methods of analysis appropriate to each review objective.

Review objective 1: quantitative studies

We included the following study designs: 1) experimental designs using random assignment to the intervention and 2) quasi-experimental designs with non-random assignment (such as regression discontinuity designs, "natural experiments," and studies in which participants self-select into the program). To be included, the studies needed to 1) collect data at baseline and endline (longitudinal) and/or cross-sectional (endline) data from treatment and comparison groups; and 2) use propensity score or other type of matching, difference-in-differences estimation, instrumental variables regression, multivariate cross-sectional regression analysis; or other forms of multivariate analysis (such as the Heckman selection model or multivariate ordinary least squares (OLS) regression analysis) that are able to correct for selection bias under specific circumstances. We included studies in which data were collected at the individual and/or group level. For studies that utilized interrupted time series, at least three data points needed to be collected before and after the intervention for the study to be included. Eligible comparison conditions were no intervention, pipeline, or

“business as usual.” We also included studies in which the outcomes of SHG members, who were member for a short amount of time, as defined by the researchers, were used as a comparison condition and/or used the time of participation in the SHG as the treatment variable. However, we were not able to include three studies that used time as a continuous explanatory variable in the meta-analysis because these studies did not allow for estimating the average impact of SHGs regardless of the time the women were members of the SHGs. We did, however, analyze the results of these studies in a narrative manner. Studies without any type of control or comparison group as outlined were excluded, including single group pre-post studies which are likely to provide biased estimates of effects due to confounding.

Review objective 2: qualitative studies

We included qualitative studies that explored empowerment from the perspectives of women participants in SHGs using the following methodologies: in-depth interviews, ethnography, participant observation, and focus groups. These studies needed to mention an underlying analytical methodology such as phenomenological analysis or grounded theory, report *actual narratives from women reported as direct quotations*, and include discussion of factors that determined women’s participation in, and benefits from, economic SHGs. Qualitative studies that did not employ the defined methodologies listed previously and that did not draw from direct quotations from female SHG participants were excluded.

3.1.5. Other study characteristics

To ensure that we included all studies since the emergence of SHGs in the early 1980s, studies were eligible which reported in any language and were conducted between 1980 and February 2014. We excluded studies that were not conducted within this time frame, with the exception of studies that were published if we had already included the working paper on which the published paper was based (Banerjee et al., 2015; De Hoop et al., 2014; Deininger & Liu, 2013).

3.2. Search methods for identification of studies

3.2.1. Electronic searches

To guide this search, we consulted an information retrieval specialist. This person is the Cochrane specialist of a research group at a large university. She gave us guidance on both search sources and search terms and built our Pubmed search strategy (below) which we used to develop all subsequent search strategies. The strategy was used to search for both qualitative and quantitative studies.

The literature search for the qualitative and quantitative studies were conducted together and this search occurred in two phases.

Phase 1: The first phase involved searching the following databases:

PubMed (<http://www.pubmed.gov>)

IndMed (<http://medind.nic.in/>)

POPLINE (<http://www.popline.org/>)

PsycINFO (<http://www.apa.org/psycinfo/>)

Index Medicus for the WHO (<http://www.globalhealthlibrary.net>)

Social Sciences Citation Index (<http://thomsonreuters.com/social-sciences-citation-index/>)

International Bibliography of the Social Sciences
(www.lse.ac.uk/collections/IBSS/about/keyFacts.htm)
British Library of Development Studies (BLDS) (<http://blds.ids.ac.uk/>)
Joint libraries of WB and IMF (JOLIS) (<http://external.worldbankimflib.org/external.htm>)
3ie Database of Impact Evaluations (<http://www.3ieimpact.org/evidence/impact-evaluations/>)
Econlit (<https://www.aeaweb.org/econlit/>)
Global Health (CABI) (<http://www.cabi.org/publishing-products/online-information-resources/global-health/>)
Africabib (<http://www.africabib.org/>)

Phase 2: Phase two consisted of reviewing reference lists of included studies and searching through studies that cited the included studies for additional resources, conducting supplemental keyword searches using identified program names and locations, and contacting key experts through an online survey for additional information.

In the second phase of the search, we also conducted a supplemental keyword search in Google.com based on leads generated by the search described above. For example, if a search identified an article mentioning (but not evaluating) a self-help group program through an MFI institution in the Philippines called *Tulay sa Pag-unlad, Inc. (TSPI)*, a search of Google.com and Google.scholar used a search of “*Tulay sa Pag-unlad Inc*” and several keywords to determine whether there was additional information on the program that might include evaluation information relevant to the analysis.

When we encountered studies that were not in English, we reviewed the English translation of abstracts that were available. We did not encounter any studies that did not have abstracts available in English. No non-English studies that had English abstracts met the inclusion criteria and therefore no further translation was needed.

We also searched the gray literature for dissertations, theses, government reports, nongovernmental organization reports, and funder reports using the following search engines and dissertations and theses.

Search engines:

IDEAS/RePEc

Google Scholar

Africa-Wide

Dissertations and theses:

ProQuest (<http://www.umi.com/enUS/catalogs/databases/detail/pqdt.shtml>)

Index to Theses (<http://www.theses.com/>)

Deutsche Nationalbibliothek (<http://www.dissonline.de/>)

We reviewed the results from these additional search engines, dissertations and theses up to 100 hits ordered by relevance since we found no relevant studies when scanning titles beyond this point.

3.2.2. Other searches

We electronically searched the collections from UC Berkeley Library and Touro University California.

We hand-searched the following key journals (specifically the past two years in case they had not been indexed in databases):

Current Anthropology, Development, Development and Change, Economic Development and Cultural Change, Feminist Economics, Global Public Health, Health Care for Women International, Health Policy, Health Policy and Planning, Indian Growth and Development Review, Indian Journal of Gender Studies, International Journal of Health Planning and Management, International Journal of Sustainable Development, International Journal of Social Research Methodology, Journal of Development Effectiveness, Journal of Development Economics, Journal of International Development, Qualitative Research in Psychology, Third World Quarterly, World Development.

We searched for relevant reports from the following multilateral organizations:

African Development Bank, Asian Development Bank, Inter-American Development Bank, International Fund for Agricultural Development, United Kingdom's Department for International Development, United Nations Children's Fund, United Nations Development Fund for Women, United Nations Development Program, United Nations Fund for Population, United States Agency for International Development, World Bank, World Health Organization.

We also contacted key personnel at the following organizations and foundations to elicit additional gray or unpublished information:

AED Center for Gender Equality, African Women's Development and Communication Network (FEMNET), Asian Women's Network on Gender and Development, the Center for the Evaluation of Global Action and Ford Foundation, Global Fund for Women, GROOTS International, The Guttmacher Institute, The Hewlett Foundation, International Committee for Research on Women, Latin American Women and Habitat Network, The Packard Foundation, SEWA, UCGHI Center of Expertise on Women's Health and Empowerment, Women Deliver.

3.2.3. Search terms

The search strategy was used to search databases and was adjusted to fit the diversity of search options available for each database. After discussion and consultation with content experts and search strategists, we included general keywords for the "exposure" and the "outcome" in our search strategy. The labeling of self-help group participation as empowering had to come from the primary researchers. We believe this strategy more accurately represented the evidence base on the impact of self-help groups on empowerment and reduced misclassification bias of our outcomes because it excluded studies in which outcome indicators did not reflect empowerment according to the group and participants under study. This decision excluded studies if these studies did not include somewhere in their text the terms "empowerment," "power," or "control." Our hand-searches and key informant contributions did not produce any additional studies that did not include at least one of these words. Thus, we are confident that our search strategy did not miss any major studies that would have been included without the exclusion criteria concerned with the terms "empowerment," "power," or "control". The search strategy was based on several consultations and discussions with our information retrieval specialist. Truncated terms and stem-words were also used where appropriate as shown in the example below.

An example of our search strategy that was used to search the PubMed database is as follows:

Search	Query	Items Found
#5	Search ((#1) AND #2) AND #3 Filters: Publication date from 1980/01/01 to 2013/12/31	1741
#4	Search ((#1) AND #2) AND #3	1811
#3	Search “women’s self-help”[tiab] OR “women’s cooperative*”[tiab] OR “self-help group*”[tiab] OR “self help group*”[tiab] OR “support group*”[tiab] OR “lending group*”[tiab] OR “advocacy group*”[tiab] OR “micro finance”[tiab] OR “micro credit”[tiab] OR “microfinance”[tiab] OR “microcredit”[tiab] OR “income generation group*”[tiab] OR “microenterprise group*”[tiab] OR sangha[tiab] OR “Self-Help Groups”[Mesh] OR (women*[tiab] AND (finance*[tiab] OR economic*[tiab])) OR (“Women”[Mesh] AND (“Financing, Organized”[Mesh] OR “Economics”[Mesh]))	29946
#2	Search “women’s empowerment”[tiab] OR “empower*”[tiab] OR “girl’s empowerment”[tiab] OR “empowering”[tiab] OR “power”[tiab] OR “control”[tiab] OR “Power (Psychology)”[Mesh]	1743835
#1	Search (“developing country” [tiab] OR “developing countries” [tiab] OR “developing nation”[tiab] OR “developing nations”[tiab] OR “developing population”[tiab] OR “developing populations”[tiab] OR “developing world”[tiab] OR “less developed country”[tiab] OR “less developed countries”[tiab] OR “less developed nation”[tiab]... [and each individual LMICs; see Appendix 2 for full list]	1139069

3.3. Data collection and analysis

3.3.1. Selection of studies

In the first stage, two team members independently reviewed titles and abstracts or executive summaries (where available) and excluded all references that were not relevant. Disagreements about inclusion were resolved through discussion. A third independent member of the team was used to resolve disagreement between the reviewers’ conclusions.

In the second stage, two team members worked independently to apply the specified inclusion criteria to the remaining full-text studies to determine whether the study should be included for analysis. Discrepancies between the two reviewers’ assessments were reviewed by a senior team member for a decision.

The full text of each study was preliminarily assessed for full-text review. These studies were retrieved and read in detail. They were screened again by four different reviewers.

3.3.2. Data extraction and management

Two team members working independently extracted information from each quantitative or qualitative study included in the review. Both team members used a pre-piloted data extraction form and the data were summarized in a table. Disagreements in coding were resolved through discussion. Study-, group-, outcome-, and effect-level data extraction and coding forms guided the data extraction (Appendix 1. Data extraction form).

3.3.3. Assessment of risk of bias in included studies

Review objective 1: quantitative studies

Two independent reviewers assessed the quantitative studies for rigor using an adaptation of a set of criteria, developed by 3ie, to assess risk of bias in experimental and quasi-experimental studies (Hombrados & Waddington, 2012). The critical appraisal tool assessed the likely risk of the following biases:

1. Selection bias and confounding, based on quality of attribution methods (mechanisms of assignment/identification), and assessment of group equivalence
2. Performance bias, based on the extent of spillovers to women in comparison groups
3. Outcome and analysis reporting biases
4. Other biases, including
 - a. Detection bias and placebo effects
 - b. Motivation and courtesy biases (Hawthorn effect and John Henry effect)
 - c. Coherence of results
 - d. Retrospective baseline data collection
 - e. Other biases, such as strong researcher involvement in the implementation of the intervention and the use of cash transfers as a compensating mechanism to participate in an intervention

The risk of bias assessment tool can be found in Appendix 6. We judged whether a study was subject to high, medium or low risk of bias for each of these categories.

We reread studies several times if something was unclear and maximized the use of all the available information from the studies. We based our assessments on the reporting in individual papers, erring on the side of caution. For example, in those cases in which the selection of participants was not clear, we classified the study as being of high risk of selection bias. In all cases where the risk of bias was unclear we assumed this was an indication of a high risk of bias.

We reported risk of bias assessment for each included study, conducting sensitivity analyses in the meta-analysis by each risk of bias domain. For example, we conducted meta-regressions to assess whether there were either substantive or statistically significant differences between low, medium, and high risks of selection bias and confounding, performance bias, outcome and analysis reporting bias, and other biases. Based on these analyses, we then determined our preferred specification for the meta-analysis. An overview of risk of bias assessment of included effectiveness studies by risk of bias category and by category of bias can be found in Appendix 7 and Appendix 8.

Review objective 2: qualitative studies

We assessed the quality of included studies using the 9-item Critical Appraisal Skills Programme Qualitative Research Checklist (CASP, 2013), making judgments on the adequacy of stated aims, the data collection methods, the analysis, the ethical

considerations and the conclusions drawn. The full checklist can be found in Appendix 5. For each item, 2 researchers determined whether the study had adequately met the item or not and gave “yes,” “no,” or “can’t tell” responses. If researchers disagreed, they discussed the item until they reached consensus. Studies that had 0-2 “no” or “can’t tell” responses were considered low risk of bias, studies that had 3-5 “no” or “can’t tell” responses were considered medium risk of bias and studies that had 6-9 “no” or “can’t tell” responses were considered high risk of bias. An overview of risk of bias assessment of included qualitative studies by risk of bias item can be found in Appendix 9.

3.3.4. Measures of treatment effects

We extracted information from each quantitative study to allow for the estimation of standardized effect sizes across studies to the extent possible. In addition, we calculated standard errors and 95 per cent confidence intervals if the information from the studies allowed for this. We conducted the sample size calculations in a consistent way to ensure comparability across studies.

The quantitative studies in our review showed substantial variation in the way they measured empowerment, even in those cases in which the studies measured the same construct. This variation was not surprising as there is no consensus as to how to measure economic, psychological, social and/or political empowerment. As discussed in our section on outcome measures, we used an iterative approach to determine whether outcome measures should be considered part of the same measurement construct. First, we grouped outcome variables under economic, social, psychological, and political empowerment. Then we synthesized the evidence based on this grouping. Finally, we conducted additional analyses to determine whether the results are robust to excluding studies with outcome measures that might not measure the same construct as the other outcome variables.

Because the studies measured empowerment in different ways, they also used different measurement scales. Several studies used dichotomous variables to measure empowerment, whereas other studies used continuous variables or indexes to measure empowerment.

Because of the different measurement scales, we report two types of effect sizes:

- 1) Standardized mean differences (*Hedges' g*).
- 2) Odds ratios.

First, we calculated the *Hedges' g* sample-size-corrected standardized mean differences (SMDs) for continuous outcome variables, which measure the effect size in units of standard deviation of the outcome variable. Second, we calculated odds ratios (ORs) for dichotomous outcome variables. The odds ratio is the ratio of the odds of an event occurring in the group of beneficiaries to the odds of the same event occurring in the comparison group (Bland & Altman, 2000). We converted the odds ratios to log odds ratios and the log odds ratios to standardized mean differences in order to make the effect sizes for continuous and dichotomous outcome variables comparable to each other. We describe the procedure for calculating the effect sizes in more detail in Appendix 10.

We converted all effect sizes to standardized mean differences to ensure we could use studies with different measurement scales in the same analysis. We found it appropriate to use dichotomous variables and continuous variables in the same meta-analysis because, in our case, variables with different measurement scales measured the same construct.

3.3.5. *Methods for handling dependent effect sizes*

We included only one effect size per study in a single meta-analysis. In one case, information was presented about the effectiveness of the same program in South Africa in two different studies. In that instance, we chose to extract effect sizes from the study that presented the most recent information (Kim et al., 2009). A different study from Ethiopia presented two impact estimates for two different regions. For this study, we calculated a pooled summary effect size using a random effect meta-analysis that included the two studies to prevent bias from dependency across the two studies. We used a random effect model because the two regions in Ethiopia can be regarded as two different contexts (Desai & Tarozzi, 2011). We included this summary effect size in the final meta-analysis.

Where studies reported more than one effect size based on different statistical methods we selected the effect size with the lowest risk of bias. We used this methodology for a study in India in which the authors used both propensity score matching and instrumental variable regression analysis to determine the impact of the program (De Hoop et al., 2014). A priori it was not clear which method had the lowest risk of bias. However, the effect size calculation clarified that the instrumental variable regression method did not result in valid effect sizes because predicted empowerment values fell outside the bandwidth of values from 0–1 for dichotomous variables. Although the impact estimates from the instrumental variable regression analysis study might have presented qualitatively interesting findings, the instrumental variable linear probability model did not show unbiased impact estimates. Hence, the risk of bias of the effect size was high. Therefore, we chose to use the impact estimates from the propensity score matching model for this study because we considered these impact estimates as medium risk of bias.

Other studies presented several impact estimates for different variables that could be argued to measure the same construct. In those cases, we chose to use either the variable that we considered the best approximation of the construct or a sample-size weighted average to measure a “synthetic effect size.” For example, in the study of Kim et al. (2009), we constructed a sample-size weighted average by estimating the average impact on self-confidence and financial confidence for psychological empowerment and on the challenging of gender norms and autonomy in decision making for social empowerment. In these cases, we used the average values of the standard errors (without weighing for the sample size) to estimate the pooled standard deviation. Similarly, for the study by De Hoop et al. (2014), we chose to calculate a sample-size weighted average for social empowerment by averaging the effects on the women’s autonomy to go to the market without their husbands’ permission and the women’s autonomy to go to the doctor without their husbands’ permission.

3.3.6. *Unit of analysis issues*

Where the standard error did not take clustering of outcomes into account in the estimation of standard errors (that is, where the outcome variables were likely to be clustered at a higher level of aggregation than the individual or household level but this was not taken into consideration in the estimation of the standard errors and confidence intervals), we used adjusted standard errors. For these studies with a risk of unit of analysis error, we applied corrections to the standard errors and confidence intervals using the variance inflation factor (Higgins & Green, 2011):

$$SE_{corrected} = SE_{uncorrected} \times \sqrt{(1 + (m - 1) \times ICC)}$$

Here, m is the number of observations per cluster and ICC is the intracluster correlation coefficient.

For the ICC, we used estimated ICCs for empowerment outcomes from a primary study in Odisha, India, that was also included in the systematic review (De Hoop et al., 2014). These ICCs were likely to be similar to ICCs in other studies, taking into consideration the large number of studies from India that we included in our systematic review. We were able to obtain the original data from the study in Odisha because one of our co-authors was also an author for this primary study (De Hoop et al., 2014). From the study in Odisha, we estimated an average ICC of 0.057 for empowerment outcomes (0.053 for social empowerment, 0.068 for psychological empowerment, 0.017 for economic empowerment, and 0.088 for measures of intimate partner violence). We used the average value of the ICC of 0.057 for the correction of the standard errors of political empowerment outcomes for which we did not have an estimate of the ICC. The other ICCs were used for the calculation of standard errors of intimate partner violence and social, psychological, and economic empowerment, respectively. If information about cluster size was not reported, we estimated the cluster size by dividing the total number of participants in each analysis (or the total number of participants if former not available) by the number of clusters. We applied this methodology to correct standard errors for 9 included studies (Ahmed, 2005; Mahmud, 1994; Nessa et al., 2012; Osmani, 2007; Rosenberg et al., 2011; Sherman et al., 2010; Steel et al., 1998; Swendeman et al., 2009; Kim et al., 2009).

3.3.7. Dealing with missing data

If the necessary data to calculate effect sizes were not available in the included studies, we attempted to contact the authors of the studies. In those cases in which we were not able to retrieve the missing data, we extracted or imputed effect sizes and associated standard errors based on commonly reported statistics such as the t or F statistic or p or z - values using David Wilson's practical meta-analysis effect-size calculator. Where studies did not report sample sizes for the treatment and the control or comparison group, we assumed equal sample sizes across the groups.

We faced several challenges with missing data in the calculation of effect sizes. First, the majority of studies that had a dichotomous dependent variable used a linear probability model rather than a logit or probit regression to estimate the effectiveness of self-help groups. Fortunately, empirically there are not many differences in marginal effects between linear probability models and nonlinear logit and probit models (Angrist & Pischke, 2009), which allowed us to estimate odds ratios under the assumption of linearity in the estimation of the standardized effect sizes. We applied this methodology to calculate effect sizes from linear probability models for dichotomous outcome variables for several studies (De Hoop et al., 2014; Desai & Tarozzi, 2011; Desai & Joshi, 2012).

Furthermore, a number of studies did not report the standard deviation of a dichotomous outcome variable, but did report the full distribution of these variables. We estimated the variance and standard deviation of these outcome variables based on the full distribution of the dichotomous outcome variables. Thus, in cases where studies reported sample sizes and the proportion of events and non-events in the sample was available we calculated the standard deviation and the effect size based on information about the sample sizes and the proportion of events and non-events.

We also included an effect size from an ordered probit regression model under the assumption that the effect size would be approximately the same if the authors had used an ordinary least squares regression model. We assumed that the point estimate from the ordered probit model would give a good estimate of the mean difference.⁴

In addition, a number of studies did not report the standard deviation of a dichotomous outcome variable but did report the full distribution of these variables. We were able to estimate the variance and standard deviation of outcome variables for which the standard deviation was not reported but for which the full distribution was reported. One study reported impact estimates using propensity score matching, but estimating the effect size in the absence of information about the standard deviation was not feasible (Deininger & Liu, 2013). In that specific case of a study from India, we imputed the standard deviation for the dichotomous outcome variables by replacing the missing standard deviations with standard deviations from similar outcome variables that were used in other studies in India (Banerjee et al., 2015; De Hoop et al., 2014).

In the absence of standard errors for the regression analysis, we also estimated the standard error of the regression analysis using the degree to which the results were statistically significant, with stars representing the significance level for one study (Mahmud, 1994).

Following all the conversions, we were able to increase the number of studies in the meta-analysis to 16 in total.

We were not able to include all studies in the meta-analysis. Two studies only demonstrated whether results were significant without the associated point estimates and standard errors. These studies did also not report t-statistics or p-values so we were not able to estimate effect sizes (Husain, Mukherjee & Dutta, 2010; Mukherjee & Kundu, 2012). One other study showed separate time-trends for latent outcome variables of the treatment and comparison group (Bali Swain & Wallentin, 2009). But these time trends alone did not allow us to extract effect sizes from the study, also because the latent variables were constructed separately for the treatment and the comparison group. The latter raises significant concerns with respect to the validity of the results. Finally, there were four studies that did not assess the impact of SHG membership but did assess the relationship between the time women were members of self-help groups (for example, in months) and women's empowerment (Coleman, 2002; Garikipati, 2008; Garikipati, 2012; Holvoet, 2005). These studies did not allow for the estimation of the average impact of women's self-help groups on women's empowerment. Nonetheless, we discuss the results of the studies narratively in our quantitative synthesis.

In those cases in which we were not able to calculate the effect size, we contacted the authors with a request for the necessary information to calculate the effect size.

3.3.8. Data synthesis

We conducted an integrated mixed-methods review in order to benefit from data generated through both quantitative and qualitative research and to enhance the review's utility and impact for policymakers. An integrated review has three stages: 1) a synthesis of quantitative effects, 2) a synthesis of relevant qualitative evidence, and 3) a synthesis of

⁴ Results were not sensitive to exclusion of this study.

both summaries that “goes beyond” the primary studies and generates new interpretations or hypotheses (Harden, 2010; Thomas et al., 2004). We conducted a meta-analysis with the data extracted from quantitative studies, and used meta-synthesis methods to synthesize the textual data extracted from the qualitative studies. We then integrated the findings from the qualitative synthesis with those from the quantitative studies to develop a framework for assessing how economic self-help groups might impact women’s empowerment.

3.3.9. Quantitative synthesis

For our quantitative synthesis (review objective 1), we statistically combined the effect sizes and associated standard errors from 23 quantitative studies that assessed the impact of self-help group programs on women’s empowerment. We only combined studies that focused on empowerment indicators that could be considered sufficiently similar. Hence, we conducted a separate meta-analysis for studies that focused on economic empowerment, social empowerment, psychological empowerment, and political empowerment, respectively. We believe these different empowerment indicators can be considered different constructs, so we did not consider it appropriate to combine these empowerment indicators in one meta-analysis.

We used inverse-variance weighted random-effects meta-analysis and used established statistical techniques to analyze heterogeneity. We used random-effects instead of fixed-effect analysis in order to allow for contextual and methodological heterogeneity in the effect sizes.

With respect to spillovers, we were unfortunately not able to report and synthesize effect sizes separately for women’s self-help group participants and neighboring women who might indirectly benefit from the intervention. None of the included studies separately reported these effect sizes.

3.3.10. Assessment of heterogeneity

We explored heterogeneity across studies with an emphasis on social and economic empowerment using I-squared and Q as well as tau-squared and the visualization of the forest plots (Borenstein et al., 2009). The results suggested there was considerable heterogeneity in the effect sizes, although less so for impacts on economic empowerment. This result was not surprising, since a substantial number of existing studies argue there is significant heterogeneity in the effectiveness of community-based programs, such as women’s self-help group interventions. This heterogeneity could be related to several contextual characteristics, such as diverging gender norms across contexts and differences in the capacity to implement community-based programs (for example, De Hoop, 2012; Mansuri & Rao, 2004; Woolcock, 2013).

It was not possible to explore heterogeneity in the impact of self-help groups on political and psychological empowerment. The number of studies focusing on these indicators was not sufficient for a reliable assessment of the heterogeneity in the impact estimates, either with a meta-analysis or with a narrative synthesis.

3.3.11. Investigation of heterogeneous effects for subgroups

We also investigated factors explaining heterogeneity by using inverse-variance weighted meta-regressions and stratified meta-analysis according to contextual and methodological moderator variables. We used two contextual moderating variables: type of intervention component; and geographic location.

We used a narrative synthesis to explore heterogeneity in the results for these subgroups because our sample of studies was relatively small. For this analysis, we integrated the findings of the qualitative analysis with the findings of the quantitative analysis to the extent possible. Hence, the potential catalysts and constraints toward the effectiveness of self-help groups that we present came from both the quantitative and the qualitative studies.

3.3.12. Sensitivity analysis

We performed an extensive sensitivity analysis for two methodological effect size moderators:

- Risk of bias status for each risk of bias category (where sufficient studies were available).
- Study design (RCTs vs. quasi-experimental studies).

We used an iterative approach based on the risk of bias assessment discussed previously to determine whether studies with different evaluation designs and different outcome measures could be combined. First, we conducted stratified meta-analyses for the randomized controlled trials and quasi-experimental evaluations in our sample. Second, we conducted meta-analyses for experimental and quasi-experimental studies with low, medium, and high risk of bias, respectively. Third, we compared the effect sizes of the different analyses to determine whether studies could potentially be combined into a single meta-analysis. In those cases in which we were not certain whether we could combine studies in a single meta-analysis, we conducted several meta-regressions to make decisions about combining studies with different characteristics in one meta-analysis. We decided to combine studies in a single analysis when the meta-regression did not show significant, either substantively or statistically, differences in the effect sizes between the studies with different risks of bias. In addition, we conducted robustness checks to determine whether studies with different outcome measures that potentially measure different empowerment constructs in the same empowerment domain (economic, social, psychological, or political empowerment) could be combined with each other in a single analysis.

We decided not to conduct meta-regressions with more than one explanatory variable because of the relatively small number of studies. Instead, we chose an iterative method in which we conducted several meta-regressions one by one to determine whether the results from studies with different methodologies and different risk of bias were sufficiently similar to combine in one meta-analysis. We started with a meta-regression to determine whether studies with studies with a low or high risk of selection bias were sufficiently similar to each other. Our approach was such that when the meta-regression presented significant, either substantively or statistically, differences between studies with a low and high risk of selection bias we excluded studies with a high risk of selection bias from the analyses. But we kept the studies with a high risk of selection bias in the analyses when the result did not show substantive or statistically significant differences between studies with a high and a low risk of selection bias. Then we continued with a meta-regression to compare findings between randomized controlled trials and quasi-experimental studies with a medium risk of bias (our synthesis did not include quasi-experimental studies with a low risk of selection bias or RCTs with a high risk of selection-bias) to see if there would be a difference between RCTs and quasi-experimental studies with medium risk of selection-bias. Similarly, we excluded quasi-experimental studies with medium risk of selection bias from the analyses if the meta-regression suggested the findings of RCTs and quasi-experimental studies with a medium

risk of selection bias were significantly, either substantively or statistically, different. But we combined the studies in a single meta-analysis if the findings of RCTs and quasi-experimental studies with a medium risk of selection bias were not substantively or significantly different from each other.

We used the same approach for different risks of bias (performance bias, outcome reporting bias, and other biases) to arrive finally at a preferred specification with randomized controlled trials with low risks of bias combined with randomized controlled trials and quasi-experimental studies with a higher (medium or high) risk of bias that did not show substantively or statistically significant different effects from randomized controlled trials with a low risk of bias. But our approach was such that we only combined RCTs with quasi-experimental studies that showed similar results in one meta-analysis to account for the possibility of selection bias in quasi-experimental studies.

We used a similar approach to determine whether studies with different outcome measures to measure the same empowerment construct (economic, social, political, and psychological empowerment) could be combined with each other in one meta-analysis. For this decision we estimated meta-analysis with and without the study or studies with a different outcome measure. We excluded studies with different outcome measures from the meta-analysis or ran a separate meta-analysis if the analysis without those studies showed substantively different effects from the analysis with those studies.

In addition, we performed a sensitivity analysis to determine whether studies with different outcome measures that could potentially measure different empowerment constructs can be used in the same meta-analysis by running meta-analysis with and without studies that use different outcome variables.

We did not conduct meta-regression analysis with more than one moderator variable in our sensitivity analysis because of the relatively small number of quantitative studies in our review.

3.3.13. Assessment of publication bias

We assessed the potential for publication bias using funnel plots for impact estimates on economic and social empowerment. In addition, we conducted Egger's test. For psychological and political empowerment, our sample size was insufficient for funnel plots to be informative about the potential for publication bias. Our sample size for political and psychological empowerment was also not sufficient for determining publication bias by comparing published with non-published studies.

3.3.14. Qualitative synthesis

The qualitative synthesis (review objective 2) was based on meta-ethnographic techniques. This process was drawn from Atkins et al. (2008), Noblit and Hare (1988) and Walsh and Downe (2005). Meta-ethnography is an interpretive approach for combining the findings of qualitative research in order to provide a higher level of analysis than individual studies alone.

Our qualitative synthesis provides a summary of women's explanations of empowerment outcomes as reported in the contributing studies. The manuscripts of the included studies were first read and reread with special attention paid to themes, quotations and authors' interpretations of the quotations. Quotations from women who discussed their experiences of empowerment were then identified and labeled with respect to the topic or concept that they represented. All quotations that were labeled or coded were subsequently categorized

into empowerment themes. This process included re-reading all labels or codes and deciding which codes were important and how they related to each other. Codes that related to similar themes were clustered together into categories that were also labeled. These categories or themes are presented in the results section with example quotations as evidence, in order to deepen readers understanding of the data. We used a systematic process to select and synthesize representative quotes from women SHG members. The selection of representative quotes was an iterative process in which two researchers identified quotations and discussed emergent themes from the included studies and determined how they were related, or dissonant, through a compare-and-contrast exercise. Typically in qualitative research, authors report 1-2 example quotations but we also provide additional quotations in Appendix 12 to improve readers' sense of the raw data and to demonstrate both the variability and the similarity between studies. Reporting 1-2 example quotations may result in reporting bias due to "cherry-picking" of non-representative quotations. Unfortunately, it is not feasible to fully account for reporting bias in qualitative research. Nonetheless, our approach, in which we provide additional representative quotes, mitigates some of the concern regarding reporting bias.

In a summary table, each category is defined, two representative quotations are given and the confidence in the findings for each category was assessed based on three areas: 1) the risk of bias assessment of the contributing studies, 2) the adequacy of the data and 3) the coherence of the theme that supported the finding. The risk of bias for each of the contributing studies is reported in the summary table based on the results of the CASP checklist as described earlier. Adequacy relates to consideration of the thickness of data and the number of studies. Thick data is achieved when detailed account of participants' experiences make explicit the phenomenon of interest. This is in contrast to a thin description, which is a more superficial account. Coherence relates to the strength of the theme across settings such as countries or regions. Based on an overall assessment of methodological quality through the risk of bias, as well as the adequacy and coherence of the data, the confidence in the evidence for each category was assessed as high, moderate, or low by two researchers and if assessed differently, they discuss until consensus was reached. A rationale with details about each confidence area is given in a summary table. The process of assessment of confidence we use is in alignment with the methodology used in Bohren et al. (2015).

3.3.15. Integrating findings from quantitative and qualitative syntheses

To integrate the findings from quantitative and qualitative synthesis, we conducted the synthesis of effects along the causal chain of the theory of change (Figure 1.1) and used the findings of the qualitative synthesis to "interrogate" and/or complement the quantitative synthesis. The information from participants gathered through qualitative investigations was used to understand whether and where any causal chain links broke down. In other words, findings from the qualitative synthesis helped describe, explore, and interpret both the nature of the empowerment process and the extent to which women experienced empowerment as recommended in the policies and guidelines of the Campbell Collaboration (Campbell Systematic Reviews, 2014).

The mixed methods review allowed us to gather information using different methodologies that informed, enhanced, and supplemented each other. The findings from the integrated synthesis were used to revise and improve our theory of change. We did this by using

information extracted from the included studies and provided insights about the nature and utility of the measures used to capture empowerment. Our aim was to synthesize the evidence produced by both bodies of research to capture the state of the evidence for the impact of self-help groups on women's empowerment.

3.4. Deviations from the protocol

The review deviates from the proposed protocol in three respects.

First, we originally intended to exclude outcomes evaluating “women's control over household resources” from microcredit self-help group studies so as not to overlap with an existing Campbell review on the impact of microcredit on women's control over household resources (Vaessen et al., 2014). However, that review does not focus specifically on self-help groups, nor does it disaggregate findings for self-help group participants and non-self-help group participants. At the same time, excluding women's control over household resources would have resulted in a considerable omission of an important outcome and undermined the comprehensiveness and value added of our review. For the sake of completeness, we have, therefore, decided to include women's control over household resources as a relevant outcome measure of economic empowerment in our review.

Second, in our original study design inclusion criteria, we specified we would only include those types of quasi-experimental studies that used statistical matching, difference-in-differences estimation, instrumental variables regression, or other forms of multivariate analysis (such as Heckman's selection models) that correct for selection bias. However, we decided also to include studies that used multivariate cross-sectional regression analysis with a dummy variable for SHG participation as a treatment variable. The identification strategy to determine causal effects of these types of studies is usually not considered credible, which may result in high risk of bias. Nevertheless, Pritchett and Sandefur (2013) proposed that including these types of studies in a meta-analysis can increase the relevance of the meta-analysis because it allows for the inclusion of studies in contexts without rigorous studies regarding the specific topic. However, we protected internal validity by a strong focus on risk of bias assessment and by conducting subgroup analyses for studies with a relatively low, medium, or high risk of bias (Campbell Systematic Reviews, 2014). In these analyses, we assessed all multivariate cross-sectional regression analysis with a dummy for SHG participation as high risk of selection-bias in a meta-regression. We then compared the estimates from studies with a high risk of selection-bias with the estimates of studies with a low-or medium risk of selection-bias. Section 4 of this systematic review shows that the findings of our review are sensitive to the inclusion of multivariate cross-sectional regression analysis. Thus, we emphasize the findings of studies with a low-or medium risk of selection-bias in the interpretation of our results.

Finally, in the protocol, we proposed to provide an overall risk of bias classification for each included study. However, to align with the most recent Campbell Collaboration best practice, we avoided using an overall quality scale and instead used risk of bias assessments for specific domains, such as selection bias and confounding, performance bias, outcome and analysis reporting bias, and other biases. Evidence suggests that assessments of overall risk of bias that do not take into consideration specific domains are too dependent on the type of quality scale used and can considerably influence the interpretation of meta-analysis results (Jüni et al., 1999). This risk of randomness in the risk of bias assessment is most likely less

severe when risk of bias assessments focus on a specific domain, such as selection bias and confounding, performance bias, outcome and analysis reporting bias, or other biases.

4. Results

4.1. Results of the search

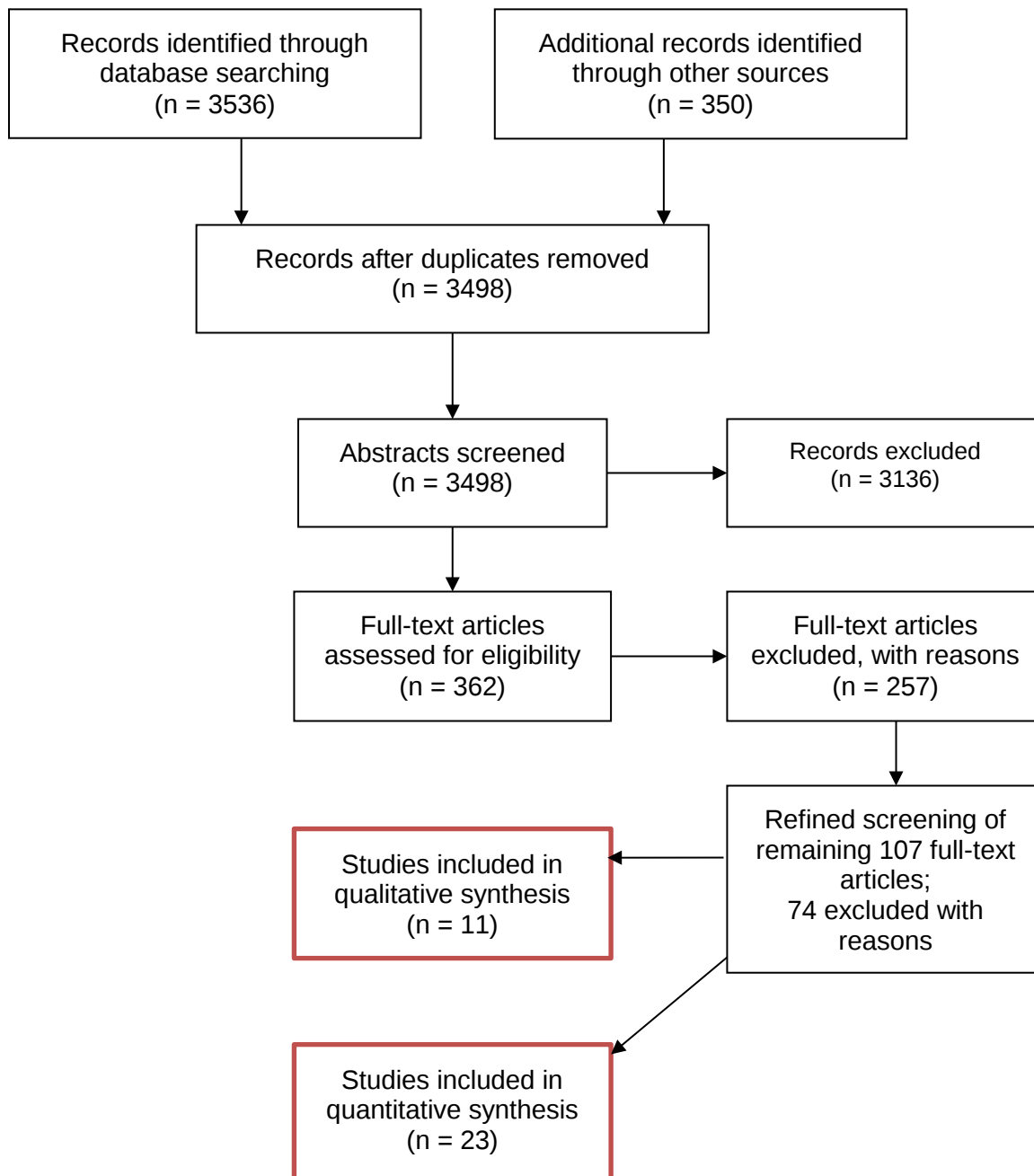
The search was conducted from March 2013–February 2014. We included a total of 23 quantitative and 11 qualitative studies in the final analysis. Figure 4.1 details the flow diagram of the filtering process used to identify the final included studies.

Initially, we reviewed 3,536 abstracts from electronic database searches and 351 abstracts from the gray literature search (see Appendix 3). Of these, we excluded 38 duplicates and 3,133 irrelevant studies. We retrieved and reviewed the full text of the remaining 365 studies using the predetermined criteria for inclusion. These studies came from database searches including library catalogues (208), hand-searches of websites (108), keyword searches (48), and author contacts (2).

Based on the full-text review of the 365 studies, we excluded 257 studies when applying the criteria. There was 93 per cent agreement among reviewers. The following were the main reasons for exclusion:

- The study did not meet our criteria of an empirical evaluation (145).
- The intervention under study did not meet our criteria of a women's economic self-help group (88).
- The evaluation design did not employ appropriate methodologies (12).
- The evaluation did not measure an empowerment outcome (7).
- The study was not focused on a self-help group in a low- or middle-income country (4).

Figure 4. 1: Study search



We reviewed the remaining 109 full-text studies (55 quantitative, 36 qualitative, 18 mixed methods) again, with specific attention paid to the methods employed. Through this process, we excluded another 74 studies because of the lack of a comparison group, a lack of quantitative estimates of impacts, a lack of a use of empowerment outcomes for quantitative studies, and a lack of data from direct observation or a lack of reporting on individual narratives for qualitative studies. Reasons for exclusion by study are reported in Appendix 4. The remaining 23 quantitative and 12 qualitative studies were included and used as the basis of the analysis that follows. Most studies were identified through database searches and came from peer-reviewed journals.

Table 4.1 summarizes the reasons for exclusion of marginal studies. The full citations of excluded marginal studies are available in the reference section and a full list of reasons for exclusion of marginal studies is available in Appendix 4. Table 4.2 summarizes sources and publication types for the included quantitative and qualitative studies.

Table 4.1: Reasons for exclusion of studies

Quantitative Studies (n=33)	
12	Study did not measure empowerment outcomes.
11	There was no comparison group.
7	Study did not evaluate a SHG.
4	There was no quantitative estimate of impact.
Qualitative Studies (n=46)	
29	Study did not evaluate the effects of a SHG.
14	Study did not report any direct quotes from participants.
3	Study did not focus on empowerment outcomes.

Table 4. 2: Sources and publication types for included studies

	Quantitative	Qualitative
Source of Included Study		
Database searches	12	6
Keyword searches	6	2
Hand-searching of organization websites	5	2
Library Catalogue	--	1
Key contact	1	--
<i>TOTAL</i>	<i>23</i>	<i>11</i>
Publication Type		
Peer-reviewed journal	15	6
Unpublished report	8	1
Book	--	1
Dissertation	--	3
<i>TOTAL</i>	<i>23</i>	<i>11</i>

4.2. Description of included studies

4.2.1 Quantitative studies (review objective 1)

The empowerment categories extracted from the quantitative studies were handled in the following way:

Economic empowerment: We included all studies measuring indicators of economic empowerment, but only meta-analyzed outcomes that focused on decision-making by women in the household. For the other indicators, we did not have a sufficient number of studies with outcome measures that were sufficiently conceptually similar to perform meta-analysis. We report the effect size findings narratively for these other indicators.

Political empowerment: We included all studies measuring indicators of political empowerment and meta-analyzed outcomes that focused on political participation, with an emphasis on voting. For the other indicators, we did not find any rigorous studies.

Social empowerment: We included all studies measuring indicators of social empowerment and meta-analyzed outcomes with an emphasis on mobility or freedom of movement and control over family size decision making jointly and separately. For the other indicators, we did not find an adequate number of studies with outcome measures that were sufficiently conceptually similar to perform meta-analysis. We report the effect size findings narratively for these other indicators.

Psychological empowerment: We included all studies measuring indicators of psychological empowerment and meta-analyzed outcomes that focused on self-confidence. For the other indicators, we did not find any studies.

All indicators were measured through household surveys, validated scales, and/or structured closed-ended questionnaires. For example, Bali Swain & Wallentin (2009) use a validated scale to measure a general empowerment index and Banerjee et al. (2015) use a normalized index score to measure economic empowerment, whereas Deininger and Liu (2013) and Holvoet (2005) use several dummy variables measured through a household survey to measure women's economic and social empowerment. De Hoop et al. (2014) use a 5-point Likert scale to measure psychological empowerment, while Kim et al. (2009) use a dummy variable indicating whether a respondent is self-confident to measure psychological empowerment. Desai and Joshi (2012) also use several dummy variables indicating women's participation in community meetings and elections to measure political empowerment.

Aggregate-level empowerment outcomes such as women's right to vote, legislation against domestic violence, inheritance law, female literacy, female child survival, and so on, were excluded from this review, also because these indicators were not clearly related to women's SHGs.

We also examined spillover effects from women's self-help group participants to nonparticipating women in the same communities. Furthermore, we examined adverse outcomes including intimate partner violence, stigma, disappointment and reduced subjective well-being.

Table 4.3 summarizes data on the SHG name, country, type of training provided, the outcome and methods used for the 23 included quantitative studies representing data from

21 SHGs, predominately based in South Asia. Of the evaluated self-help groups, 11 were implemented in India and 6 were implemented in Bangladesh. The remaining studies came from Thailand (1), South Africa (1), Ethiopia (1), and Haiti (1). Two self-help groups (one from India and one from South Africa) were discussed in two quantitative papers. One study consisted of two separate analyses for samples in two different regions in Ethiopia (Desai & Tarozzi, 2011). All of the study findings were based on analyses of self-reported survey data either based on experimental or observational designs.

Although in most cases detailed information on the intervention activities was not recorded clearly, several studies present some information about whether any training or services was offered to the SHG and the type of training offered. Ten of the self-help groups did not report any additional training or services beyond financial services (credit, loans, and savings). The remaining 11 groups offered some combination of the following: health education (4), business or entrepreneurial skills (6), awareness of women's rights (2), basic education (2), and community-development training (2). However, this list of training and supplemental activities only represents what was reported by authors.

All of the included self-help groups were initiated by local or international NGOs and community-based organizations. Four of the 20 groups were initiated by the Grameen Bank and three by the Bangladesh Rural Advancement Committee (BRAC). Only two of the groups, represented in three studies, were initiated as the intervention arm of a research study (Pronyk et al., 2006; Kim et al., 2009; Sherman et al., 2010).

The study designs and methods of analysis used in the studies were very diverse. Four studies used cluster-randomized assignment (Desai & Joshi, 2012; Desai & Tarozzi, 2011; Kim et al., 2009 (incorporating Pronyk et al., 2006); Sherman et al., 2010). The remaining studies were based on observational data using methods of counterfactual identification such as propensity score matching (PSM) (de Hoop et al., 2014), PSM combined with double-differences (Deininger & Liu, 2009) and instrumental variables analysis (Osmani, 2007; Pitt et al., 2006). Methods used to estimate treatment effects ranged from ordinary least squares regression analysis (for example, Osmani, 2007) and logistic regression (for example, Ahmed, 2005) to the calculation of risk-or odds ratios based on events/non-events (for example, Swendeman et al., 2009).

Table 4. 3: Summary of included quantitative studies

Study	SHG Name	Setting	Additional Training	Outcome	Sample	Data Collection	Study design	Analysis
Ahmed, 2005	BRAC	Bangladesh	Business Skills	(1) Intimate Partner Violence	2044 Households with currently married women in 60 villages	Survey data	Cross-sectional observational study	Logistic Regression
Banerjee et al., 2015	Spandana	Hyderabad, India (South)	None	(1) Economic Empowerment	1220 households in 52 randomly selected neighborhoods that are eligible for Spandana microfinance	Survey data	Repeated cross-section cluster-randomized controlled trial	Linear probability model (OLS)
Bali Swain & Wallentin, 2009	SHG Bank Linkage Program	India (5 states)	None	(1) Empowerment Index	1000 households in 2 representative districts in 5 Indian states that were randomly selected from SHG members, and a comparison group that is comparable in terms of socio-economic characteristics	Survey data	Observational study with panel data	Analysis of separate time trends for treatment and comparison households

Study	SHG Name	Setting	Additional Training	Outcome	Sample	Data Collection	Study design	Analysis
Coleman, 2002	Bank for Agriculture and Agricultural Cooperatives	Thailand	None	(1) Economic Empowerment	445 households in 14 villages	Survey data	Observational study with panel data	OLS regression analysis
De Hoop et al., 2014	CENDERET	Orissa, India	Business Skills, Rights Awareness	(1) Economic Empowerment (2) Social Empowerment (Mobility) (3) Psychological Empowerment (4) Intimate Partner Violence	398 households in 19 villages	Survey data	Cross-sectional observational study	Propensity score matching
Deininger and Liu, 2009	Indhira Kranthi Patham Program	Andhra Pradesh, India (South)	Community Development	(1) Economic Empowerment (2) Social Empowerment (Mobility) (3) Political Empowerment	6340 households from 659 villages	Survey and census data	Observational study with cross-sectional and recall data	Double-difference propensity score matching
Desai and Joshi, 2012	Self-Employed Women's Association (SEWA)	Rajasthan, India (North)	Business Skills , Child Care Services, Employment Training,	(1) Economic Empowerment (2) Social Empowerment (Family Size Decision Making)	3535 households from 82 villages	Survey data	Repeated cross-section cluster-randomized controlled trial	Linear probability model with block-level fixed effects

Study	SHG Name	Setting	Additional Training	Outcome	Sample	Data Collection	Study design	Analysis
			Leadership Training	(3) Political Empowerment				
Desai and Tarozzi, 2011	Amhara Development Association	Ethiopia	Family Planning Services	(1) Social Empowerment (Family Size Decision Making)	1600 households in 54 villages	Survey Data	Repeated cross-section cluster-randomized controlled trial	Linear probability model
Desai and Tarozzi, 2011	Oromia Credit and Savings Share Company, & Oromia Development Association	Ethiopia	Family Planning Services	(1) Social Empowerment (Family Size Decision Making)	1600 households in 54 villages	Survey Data	Repeated cross-section cluster-randomized controlled trial	Linear probability model
Garikipati, 2008	National Bank for Rural and Agricultural Development	Andhra Pradesh, India (South)	None	(1) Economic Empowerment	291 households in 2 villages	Survey data	Observational study with cross-sectional data	Instrumental variable regression analysis with time spent in SHG as explanatory variable
Garikipati, 2012	National Bank for Rural and Agricultural Development	Andhra Pradesh, India (South)	None	(1) Economic Empowerment	291 households in 2 villages	Survey data	Observational study with cross-sectional data	Instrumental variable regression analysis with time spent in SHG as

Study	SHG Name	Setting	Additional Training	Outcome	Sample	Data Collection	Study design	Analysis
								explanatory variable
Holvoet, 2005	IRDP & TNWDP	Tamil Nadu, India (South)	Rights Awareness, Community Development, Business Skills	(1) Economic Empowerment (2) Social Empowerment (Family Size Decision Making)	300 households in 6 blocks	Survey data	Observational study with cross-sectional data	Multinomial logit model with time spent in SHG as explanatory variable
Husain et al., 2010	Swarna Jayanti Sahari Swarojgar Yojana	West Bengal, India	None	(1) Economic Empowerment (2) Social Empowerment (Mobility)	Household data from unknown number of households from 6 municipalities	Survey Data	Observational study with cross-sectional data	Logit model with time spent in SHG as explanatory variable
Kim et al., 2009 (incorporating Pronyk et al., 2006)	Intervention with Microfinance for AIDS and Gender Equity	South Africa	Health Education (HIV prevention)	(1) Economic Empowerment (2) Social Empowerment (Mobility) (3) Psychological Empowerment (4) Intimate Partner Violence	1409 households in 12 villages	Survey Data + census Data	Cluster-randomized controlled trial with cross-sectional data	Cross-sectional comparison to determine risk ratio

Study	SHG Name	Setting	Additional Training	Outcome	Sample	Data Collection	Study design	Analysis
Mahmud, 1994	BRAC, Grameen Bank, Bangladesh Rural Development Board, Women's Entrepreneurship Development Program	Bangladesh	Health Education (family planning)	(1) Social Empowerment (Family Size Decision Making)	806 households with currently married women in 8 villages	Survey Data	Observational study with cross-sectional data	Logit model
Mukherjee & Kundu, 2012	Swarnajayanti Gram Swarojgar Yojana	West Bengal, India	None	(1) Economic Empowerment (2) Social Empowerment	500 households in 14 villages	Survey Data	Observational study with panel data	Multinomial logit model
Nessa et al., 2012	Grameen Bank, BRAC, and ASA	Bangladesh	None	(1) Economic Empowerment (2) Social Empowerment (Mobility) (3) Political Empowerment	600 households in 8 districts	Survey Data	Observational study with cross-sectional data	OLS regression model with time spent in SHG as explanatory variable
Osmani, 2007	Grameen Bank	Bangladesh	None	(1) Economic Empowerment	84 households in 4 villages	Survey Data	Observational study with cross-sectional data	Instrumental variable regression analysis

Study	SHG Name	Setting	Additional Training	Outcome	Sample	Data Collection	Study design	Analysis
Pitt et al., 2006	Grameen Bank	Bangladesh	None	(1) Economic Empowerment (2) Social Empowerment (Mobility + Family Size Decision Making) (3) Political Empowerment	1798 households from 87 villages	Survey Data	Observational study with panel data	Instrumental variable regression analysis
Rosenberg, 2011	Fondasyon Kole Zepol	Haiti	Basic Education, Business skills, Rights Awareness, Health Education	(1) Social Empowerment (Family Size Decision Making)	192 households that selected in the survey	Survey Data	Observational study with cross-sectional data	OLS regression model with dummy variable that is 1 if clients had been involved in the SHG for more than 12 months and 0 if less than 12 months
Sherman et al., 2010	Research Study John Hopkins School of Public Health	Chennai, India (South)	Health Education (HIV prevention), Business Skills	(1) Economic Empowerment (Sex Partners)	100 sex-workers in Chennai	Survey Data	Cluster-randomized controlled trial with panel data	Difference-in-difference analysis

Study	SHG Name	Setting	Additional Training	Outcome	Sample	Data Collection	Study design	Analysis
Steel et al., 1998	Save the Children & Association for Social Development	Bangladesh	None	(1) Social Empowerment (Family Size Decision Making)	6456 households in 15 villages	Survey Data	Observational study with panel data	Difference-in-difference analysis
Swendeman et al., 2009	Sonagachi Project	West Bengal, India (Central)	Health Education (HIV prevention)	(1) Economic Empowerment (Sex Workers) (2) Social Empowerment (Sex Workers) (3) Psychological Empowerment (Sex Workers)	110 sex-workers in two towns	Survey Data	Observational study with cross-sectional data	Cross-sectional comparison to determine odds ratio

4.2.2 Qualitative studies (review objective 2)

The 12 qualitative studies were also predominately from South Asia. Nine studies focused on SHGs in India. The remaining studies came from Nepal (1), Bolivia (1) and Tanzania (1). Table 4.4 describes the included qualitative studies, including the name of the SHG, the setting, the sample, the data collection, and the methods of analysis.

Most of the qualitative data were drawn from purposive or convenience samples of SHG participants through unstructured or semi-structured in-depth interviews. Two studies (Dahal, 2014; Ramachandar & Pelto, 2009) randomly selected participants. Two other studies, (Maclean, 2012; Mercer, 2002) used a case study methodology to describe how SHGs operate within a village context. Six studies (Dahal, 2014; Knowles, 2014; Kilby, 2011; Maclean, 2012; Mercer, 2002; Sahu & Singh, 2012) used focus groups in addition to individual interviews. Most of the studies did not name the specific qualitative theory behind their analysis methodology but descriptions of their analysis process indicated that most studies used some adaptation of grounded theory, content analysis or thematic analysis techniques.

We present further details in the qualitative synthesis below (Chapter 4.4).

Table 4. 4: Summary data on qualitative studies

<i>Author, Year</i>	<i>Name or Description of SHG</i>	<i>SHG main activities</i>	<i>Setting</i>	<i>Sample</i>	<i>Data Collection</i>	<i>Analysis</i>
Dahal 2014	Village Development Committees in Lamachaur	Microcredit, trainings and social awareness	Nepal	Random Sample of 40 female SHG members, and 3 SHG leaders	Focus groups and in-depth interviews	Thematic Analysis
Kabeer 2011	BRAC, Nijera Kori, Saptagram and Samata	<u>BRAC</u> : Microcredit entrepreneurial skills, literacy; <u>Nijera Kori, Saptagram and Samata</u> : Savings, activism and collective awareness raising	Bangladesh	Convenience selection of 31 women from 4 socially oriented SHGs	Loose life history approach, semi-structured interviews	Modified Grounded Theory

<i>Author, Year</i>	<i>Name or Description of SHG</i>	<i>SHG main activities</i>	<i>Setting</i>	<i>Sample</i>	<i>Data Collection</i>	<i>Analysis</i>
Kilby 2011	77 small local groups in Karnataka and Pune	Microfinance; community based management of natural resources, sustainable agriculture, human rights	South India	Women from 70 purposively selected NGO-initiated self-help groups	70 focus groups, 2 workshops and key informant interviews	Modified Grounded Theory using Ranking Exercise
Knowles 2014	Tamil Nadu Women's Association	Microfinance; community development	South India	Purposive selection of 196 female SHG members	In-depth semi-structured interviews, Structured Focus Groups, Participant Observation	Content Analysis
Kumari 2011	Gandhi Smaraka Grama Seva Kendram in Kerala	Microfinance	South India	Purposive sample from networked groups of women in one urban slum and one tribal area	Participant observation, informal chats, focus group discussions and interviews	Phenomenology
Maclean 2012	Credit with Rural Education in Luribay	Microfinance, Village banking training	Bolivia	Case study of one village banking program	28 in-depth interviews, 2 focus groups, and participation in 40 group meetings	Case Study
Mathrani 2006	Mahila Samakhya in Karnataka	Microsaving, literacy training, community development	South India	Seven purposively-selected village SHGs	Unstructured interviews with participants	Modified Grounded Theory

<i>Author, Year</i>	<i>Name or Description of SHG</i>	<i>SHG main activities</i>	<i>Setting</i>	<i>Sample</i>	<i>Data Collection</i>	<i>Analysis</i>
Mercer 2002	Chagga village women's organizations in Hai District	Cooperative income-generation activities	Tanzania	Case study of four village-based women's organizations	Group discussions, qualitative household interviews	Case Study
Pattenden 2011	Jagruthi Mahela Sanghathan in Karnataka	Income Diversification and Trainings (gender violence, discrimination, health, rights, agriculture)	South India	All the members of three purposively selected, scheduled caste women's associations in 3 villages	Two rounds of semi-structured interviews	Modified Grounded Theory
Ramachandrar 2009	Family Planning Association of India in Bellary	Microfinance and Family Planning, training (gender issues, credit management, leadership, income generating activities)	South India	Random selection of 25 SHGs from 50 total groups within one organization	In-depth semi-structured interviews	Modified Grounded Theory
Sahu 2012	Madagadipet Self Help Groups	Microfinance	South India	Convenience sample of female SHG members from 6 different groups	6 Focus Group Discussions	Content Analysis

4.3. Critical appraisal of included studies

4.3.1 Risk of bias of quantitative studies

We relied on a risk of bias tool with 71 criteria that were related to selection bias and confounding, performance bias, outcome and analysis reporting biases, and other biases. The complete tool and a detailed assessment of the risk of bias of each individual quantitative study can be found in Appendices 7.6 and 7.8.

Figure 4.2 shows that only three of the 23 quantitative studies were rated as having a low risk of selection bias. Each of these studies was a cluster-randomized controlled trial with a sufficient sample size to ensure equivalence in observable and unobservable characteristics across the treatment and the control group. RCTs with a small sample size were rated as having a medium risk of selection bias because the studies usually did not show sufficient evidence that there was equivalence in observable characteristics. In addition, quasi-experimental studies were usually not convincing in their claims that selection bias was no longer an issue after controlling for observable characteristics with statistical tools, such as propensity score matching and multivariate regression analysis. We rated studies that used propensity score matching with a large number of plausibly exogenous control variables as having a medium risk of selection bias and studies that used multivariate regression analysis as having a high risk of selection bias.

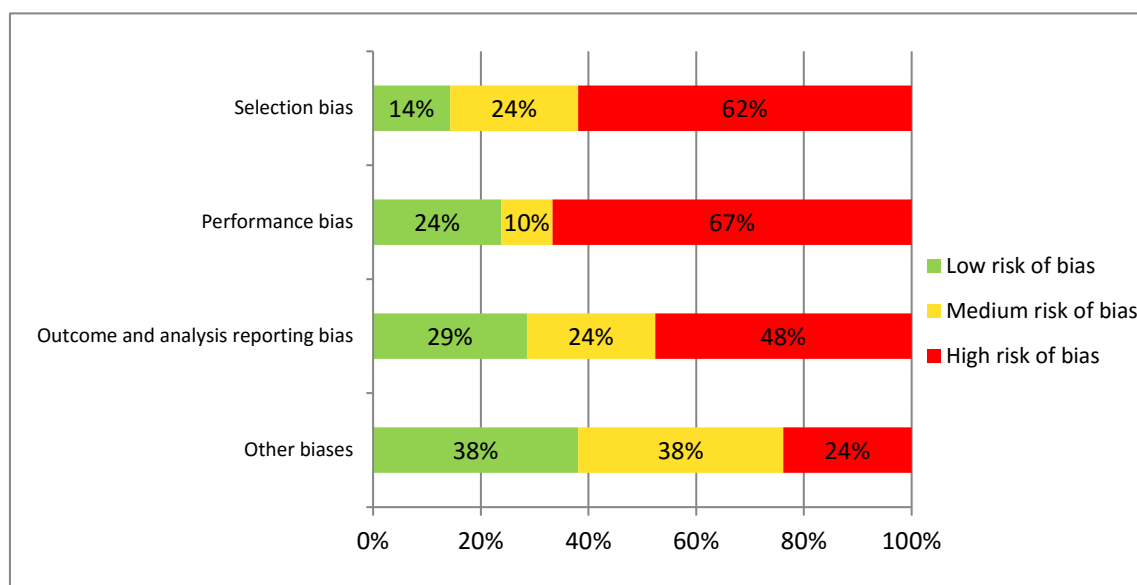
Of the 23 quantitative studies, five studies were rated as having a low risk of performance bias. These studies usually had a control or comparison group that was not in direct contact with the beneficiaries of the intervention to ensure the control or comparison group was not contaminated by the intervention or the adoption of practices by beneficiaries of the intervention as a result of their SHG membership. Studies that included a comparison group that was in direct contact with the beneficiaries but that took measures in their analysis or sampling strategy to consider this were rated as having a medium risk of performance bias. For example, Banerjee et al. (2015) acknowledged that the control group was contaminated by other microfinance services similar to the intervention they evaluated. However, the authors also demonstrated that the uptake of microcredit by beneficiaries was significantly higher in the treatment villages. Hence, performance bias could be rated as medium in this specific study. Other studies that included a comparison group that was in close contact with the beneficiaries were rated as having a high risk of performance bias.

Of the 23 included studies, six studies were rated as having a low risk of outcome and analysis reporting bias. These studies did not show signs of inconsistent reporting or unusual types of analyses. Several other studies were labeled as medium risk of outcome and analysis reporting bias because of unclear explanation of the outcome variables or the use of potentially flawed analyses. For example, we rated studies that used potentially endogenous variables as explanatory variables as having a medium risk of outcome and analysis reporting bias. In these cases the outcome equations were potentially incorrectly specified. Finally, several studies did only show tables for outcome variables that were significantly affected by self-help groups and not for outcome variables that were not significantly affected. We labeled these studies as having a high risk of outcome and analysis reporting bias because of the potential for publication bias. We also labeled studies that used an explanatory variable with the amount of time that respondents were members of SHGs as having a high risk of outcome and analysis reporting bias. Such explanatory variables increase the risk of bias due to a lack of accounting for potential nonlinearities in the impact estimates of SHGs.

Finally, of the 23 included studies, eight were rated as having a low risk of other bias. These studies did not show any other potential biases. But another 38 per cent of the studies were rated as having a medium risk of potential bias, for example, because studies did not explain well whether authors took measures to mitigate concerns regarding the measurement of potentially sensitive outcome variables, such as domestic violence. Studies with a high risk of other biases included studies that relied extensively on recall data for outcome variables, which raised the likelihood of social desirability bias. For example, SHG members may have had the perception that enumerators would like to hear that SHG membership has resulted in improvements in autonomy. Under such circumstances, the respondents might have an incentive to underestimate their level of autonomy before the start of their SHG membership and to overestimate their level of autonomy after the start of the SHG membership.

There was almost complete agreement between the two reviewers in assessments of the risk of selection and performance bias, but initially there were more disagreements about the risk of outcome and analysis reporting biases and other biases. In first instance, the reviewers disagreed about the risk of selection bias and confounding for one of the 23 included studies (Desai & Tarozzi, 2011), risk of performance bias for two of the 23 included studies (Holvoet, 2005; Sherman et al., 2010), risk of outcome and analysis reporting bias for seven of the 23 included studies (Ahmed, 2005; De Hoop et al., 2014; Sherman et al., 2010; Rosenberg et al., 2011; Swendeman et al., 2009; Pitt et al., 2006; Nessa et al., 2012) and other biases for 11 of the 23 included studies (Coleman, 2002; Banerjee et al., 2015; Desai and Joshi, 2012; Garikipati, 2008; Garikipati, 2012; Kim et al., 2009; Pronyk et al., 2006; Mukherjee and Kundu, 2012; Rosenberg et al., 2011; Osmani, 2007; Steele et al., 1998). However, in all cases where there was no immediate agreement, the reviewers reached agreement about the risk of bias assessment through consensus.

Figure 4. 2: Risk of bias assessment of quantitative studies



4.3.2 Quality of qualitative studies

The appraisals of the qualitative studies are summarized in Figure 4.3 and assessments by study are included in Appendix 9.⁵ The nine-question tool aimed to determine whether a study was valid if the results were reported adequately and if the findings would be helpful locally. The nine studies were considered valuable based on responses to two screening questions and seven assessment questions. There was almost complete agreement between the two researcher assessors. In two cases associated with consideration of ethical issues and one case associated with the relationship between the researcher and the participants, one researcher felt that she could not tell whether a criterion was met, whereas the other researcher was able to identify the information to answer the criteria (Pattenden, 2011; Ramachandar & Peltó, 2009; Kumari, 2011).

Figure 4. 3: Summary of quality appraisal of qualitative studies

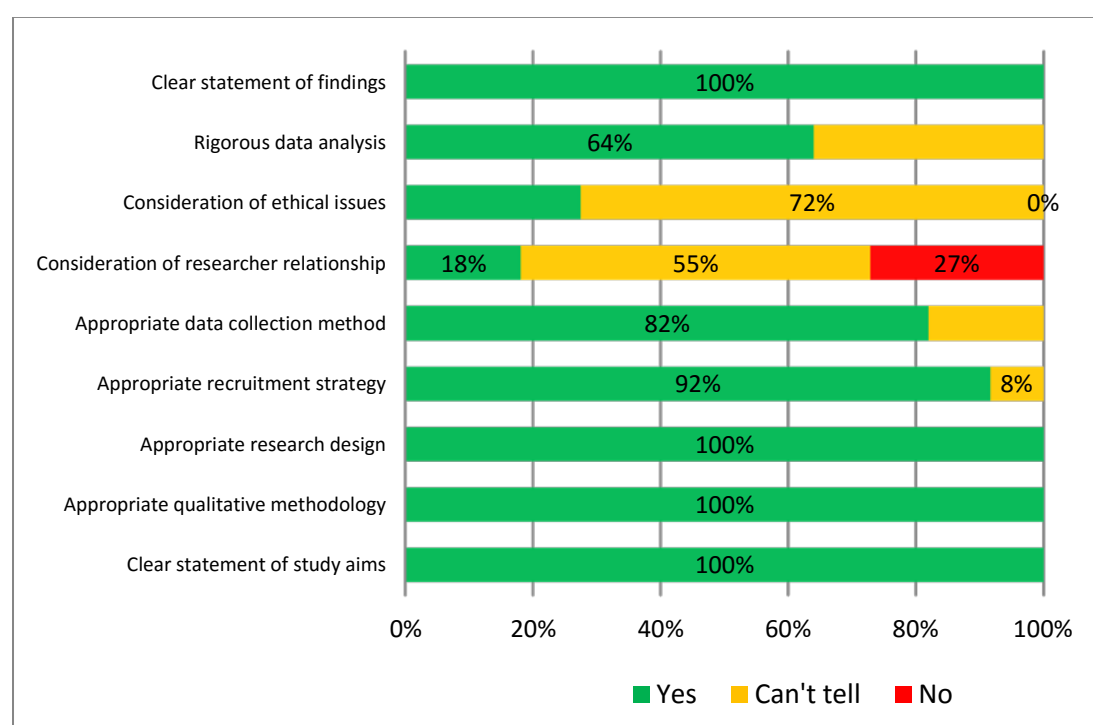


Figure 4.3 shows that there are important concerns regarding several of the quality criteria for the included qualitative studies, although all included qualitative studies had a clear statement of the study aims, appropriately used qualitative methodology, had an appropriate research design, and reported clear statements of their findings. Studies that received a “can’t tell” or “no” did so for several main reasons. With respect to the recruitment strategy, authors did not always explain how the participants were selected and why this selection could be considered the most appropriate sampling strategy for the study. There was also not sufficient explanation of the recruitment process such as who chose to participate and who declined. With respect to data collection, authors did not adequately justify why they had chosen one method over another. Few authors described their data collection tools such as interview guides or their data format such as tape recordings or handwritten notes. No

⁵ Details of the quality appraisal assessment criteria are in Appendix 5.

author mentioned data saturation as a reason for stopping recruitment. Most authors did not report information about the researcher-participant relationship and did not examine the potential bias and influence they introduced during all aspects of the study. In addition, very few authors described whether and how ethical standards were maintained (such as informed consent). The authors also did not discuss any ethical issues that the study raised. Finally, many studies lacked an in-depth description of the data analysis process both in terms of the methodology used and how the analysis was carried out.

4.4. Synthesis of quantitative studies

This section presents results of meta-analysis of the effects of women's self-help groups on women's economic, social, psychological, and political empowerment and intimate partner violence (review objective 1). In addition to the preferred specification for economic, social, psychological, and political empowerment and intimate partner violence, we also present an extensive sensitivity analysis with separate impact estimates for studies with high, medium, and low risk of bias, and randomized controlled trials and quasi-experimental evaluations. Further, we analyze heterogeneity by comparing effect sizes across geographic contexts, although our sample size only permitted a narrative analysis of the differences across geographic contexts. We also present a narrative analysis to determine the separate effects of different components of self-help groups, such as microcredit, microsavings, and training. Finally, we present a narrative analysis to determine differences in effect sizes between studies within the same empowerment domain that have different outcome measures and might thus measure different empowerment constructs.

4.4.1 Economic Empowerment

Of the 23 included quantitative studies, ten included an impact estimate on women's economic empowerment that we were able to include in our meta-analysis, and eight included an impact estimate on women's economic empowerment but did not allow for determining the effect size of the intervention. We summarize the measurement of economic empowerment and the feasibility to include studies in the meta-analysis in Table 4.5.

Table 4. 5: Measurement of women's economic empowerment

Study	Definition of Variable	Scale	Included in Meta-Analysis?
Bali Swain & Wallentin (2009)	General index of women's empowerment.	Normalized score from 0-1	No; not able to estimate effect size
Banerjee et al. (2015)	Normalized index score that includes variables that measure the decision-making power of the female respondent in the household.	Normalized score from 0-1	Yes
Coleman (1999)	Several variables that emphasize the female ownership of assets.	Several binary variables	No; not able to estimate effect size
De Hoop et al., (2014)	Dummy variable that is 1 for women who make decisions about food expenditures.	Binary	Yes
Deininger and Liu (2013)	Dummy variable that is 1 for women who are able to save individually.	Binary	Yes; after imputing the standard deviation
Desai and Joshi (2012)	Several dummy variables associated with women's decision-making power about schooling and health expenditures.	Several binary variables	Yes
Garikipati (2008) & Garikipati (2012)	Women's labor supply.	Continuous variable	No; not able to estimate effect size

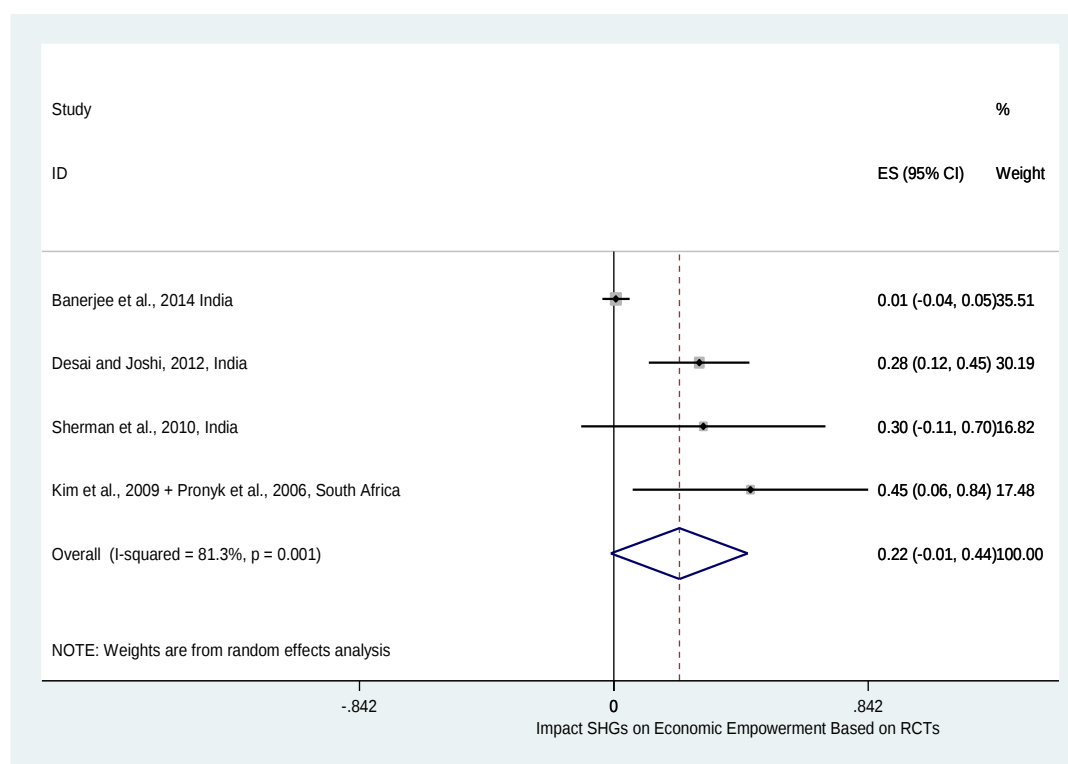
Study	Definition of Variable	Scale	Included in Meta-Analysis?
Holvoet (2005)	Several dummy variables associated with women's decision-making power in economic and non-economic domains.	Several binary variables	No; not able to estimate effect size
Husain et al. (2010)	Several variables associated with women's decision-making power in the economic domain, which are 0 if the woman has no decision-making power, 0.5 if there is joint decision-making and 1 if the woman is the sole decision-maker.	Aggregate score of categorical variables	No; not able to estimate effect size
Kim et al. (2009) & Pronyk et al. (2006)	Dummy variable that is 1 if the woman believes her contribution to the household is positive	Binary variable	Yes
Mukherjee & Kundu (2012)	Several dummy variables associated with women's decision-making power about household expenditures.	Index of binary variables	No; not able to estimate effect size
Nessa et al. (2012)	Categorical variable associated with the economic decision-making power of the woman in the household.	Binary variable	Yes
Osmani (2007)	Categorical variable that measures the perception of the woman on how well she would be able to take care of herself.	Ordered categorical variable	Yes
Pitt et al. (2006)	Several dummy variables that are associated with women's decision-making power about household expenditures, access to funds, and borrowing money.	Index of binary variables	Yes
Sherman et al. (2010)	Self-reported number of sex-exchange partners.	Continuous variable	Yes
Steel et al. (1998)	Several dummy variables associated with women's decision-making power with respect to medical expenditures, borrowing, and housing repairs.	Several binary variables	No; not able to estimate effect size
Swendeman et al. (2009)	Dummy variables related to decision-making power of female sex workers.	Several binary variables	Yes

The table demonstrates that women's empowerment was measured in different ways across studies. However, with a few exceptions, women's economic empowerment was reflected in women's bargaining power or decision-making power. We were not able to include the few studies that do not measure women's bargaining power but another component of women's economic empowerment in the meta-analysis because we were not able to calculate effect sizes for these specific studies. We discuss the results of these studies in a narrative synthesis. The measurement of women's bargaining power was mostly associated with decisions about expenditures and borrowing, but for the specific case of sex workers bargaining power was also associated with decision-making power about the number of clients for the sex worker.

The measurement of women's bargaining power might thus measure a different construct for sex workers. Therefore, we conducted meta-analyses with and without studies that measure women's bargaining power for sex workers. In addition, we also conducted a meta-analysis without the study of Deininger and Liu (2013) who emphasize women's ability to save individually. Although this concept might be related to women's bargaining power, women's ability to save individually could also be considered a different construct.

Figure 4.4 presents the forest plot with the results of the meta-analysis of randomized controlled trials. From the analysis, it appears that women's self-help groups have an average positive effect of 0.22 standard deviations on women's economic empowerment (SMD=0.22, 95% confidence interval (CI)=-0.01, 0.44; evidence from 4 studies), but one which is not statistically significant at the 95 per cent level. The meta-analysis also suggests strong heterogeneity in the impact estimates of women's self-help groups on women's economic empowerment. Observed heterogeneity in effect sizes ranged from 0.01 to 0.45 standard deviations and statistical tests suggest there is support that this heterogeneity is real rather than due to random sampling error ($Q=16$, $\text{Tau-sq}=0.04$, $I\text{-sq}=81\%$). However, we are not able to interpret $I\text{-squared}$ as an absolute indicator of heterogeneity (Borenstein et al., 2009; Higgins, 2011), and the estimate of the variance component tau-squared is low, suggesting the level of between-study heterogeneity may be limited. We should be careful in interpreting these results, however, as these tests are not always appropriate for a small number of studies (ibid.).

Figure 4. 4: The effects of women's self-help groups on women's economic empowerment (randomized controlled trials)

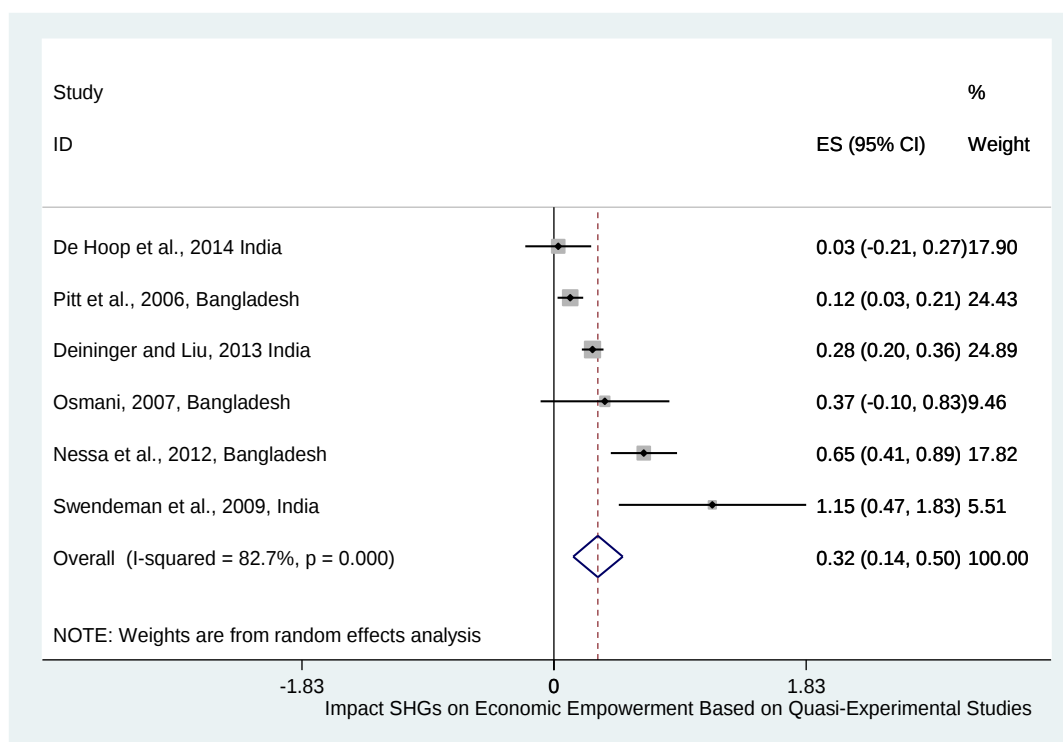


There are various differences in the implementation, context and risk of bias of the RCTs. First, there was heterogeneity in the types of self-help groups that were evaluated using randomized controlled trials. For example, the study by Banerjee et al. (2015) focused on a self-help group intervention without a training component. And the study of Sherman et al. (2010) assessed the impact of a women's self-help group program on the economic empowerment of female sex workers. Arguably, the included studies were not fully comparable to each other and this needed to be taken into consideration in a sensitivity analysis. We illustrate this by a meta-regression that demonstrated that the estimated effect sizes on economic empowerment of RCTs of interventions with a training component were substantively and statistically significantly higher (SMD=0.31, 95% CI=-0.16, 0.45; Q=0.6, Tau-sq=0.00, I-sq=0%; evidence from 3 studies) than the effect size of the study by Banerjee et al. (2015). The changes in the confidence interval and the reductions in the indicators to measure heterogeneity after excluding the studies without a training component also suggest that SHG programs with training have substantively higher effect sizes on women's bargaining power than SHG programs without a training component and that heterogeneity is mostly caused by including studies with a training component. The effect size of the study of Sherman et al. (2010) with an emphasis on sex workers is also not substantively different from the effect sizes of other interventions with a training component. Thus, excluding the study of Sherman et al. (2010), which might potentially measure a different empowerment construct, does not change the interpretation of the results. The average effect size of SHGs on women's economic empowerment is 0.20 SMD when we exclude studies with an emphasis on sex workers (SMD=0.20, 95% CI=-0.05, 0.46; Q=14, Tau-sq=0.04, I-sq=86% ; evidence from 3 studies) and 0.31 SMD when we exclude studies with an emphasis on sex workers and studies without a training component (SMD=0.31, 95% CI =0.16, 0.46; Q=0.62, Tau-sq=0.00, I-sq=0% ; evidence from 2 studies).

The effect sizes of each of the studies included in Figure 4.4 were all potentially subject to various biases despite the random allocation of the intervention. Both the study of Sherman et al. (2010) and Kim et al. (2009) were rated as having a medium risk of selection bias due to the small sample size of these studies. Furthermore, the study of Banerjee et al. (2015) was rated as having a medium risk of performance bias because of contamination of the control group by various other microfinance initiatives. In addition, the study of Sherman et al. (2010) was rated as having a high risk of performance bias because the control group lives in the same locality as the beneficiaries of the intervention, which may result in spillovers. Meta-regressions did not suggest statistically significant differences in effect sizes between RCTs that were rated as having differential risks of bias. Nonetheless, the evidence for heterogeneity suggested that we were not able to derive strong conclusions about the effects of women's self-help groups on women's economic empowerment based on these studies alone. Furthermore, statistical heterogeneity in the estimates suggested that it might be beneficial to include additional studies with a higher degree of precision.

We conducted a separate meta-analysis of the effects of women's self-help groups on women's economic empowerment based on quasi-experimental evaluations (Figure 4.5). From the analysis, it appears that women's self-help groups have a positive effect on women's economic empowerment, which is statistically significant at the 95 per cent level (SMD=0.32, 95% CI=0.14, 0.50; evidence from 6 studies). Again, the meta-analysis suggested strong heterogeneity. Effect sizes of the studies ranged between 0.03 and 1.15 standard deviations, while statistical heterogeneity tests suggested that a substantial percentage of the observed heterogeneity in the effect size is real rather than random sampling error ($Q=29$, $I\text{-}sq=83\%$), albeit with a small estimated variance component ($\tau\text{-}sq=0.03$).

Figure 4. 5: The effects of women's self-help groups on women's economic empowerment (quasi-experimental evaluations)



Additional analysis suggested that the heterogeneity in the impact estimates could be partly explained by the inclusion of quasi-experimental studies with a high risk of selection-bias. Meta-analyses of quasi-experimental studies with a medium and high risk of selection-bias indicated that the impact estimate of studies with a high risk of selection-bias is notably higher than the impact estimate of studies with a medium risk of selection-bias (Figure 11.1 and 11.2 in Appendix 11). The meta-analyses indicated that the impact estimate of quasi-experimental studies with a high risk of selection bias was on average 0.65 standard deviations (SMD=0.65, 95% CI=0.33, 0.98; Q=29, Tau-sq=0.04, I-sq=42%; evidence from 3 studies), which is approximately three times as high as the effect size for RCTs (0.22 standard deviations). The average impact estimate of studies with a medium risk of selection bias of 0.17 standard deviations (SMD=0.17, 95% CI=0.03, 0.34; Q=9, Tau-sq=0.01, I-sq=78%; evidence from 3 studies) is much closer to the impact estimate of randomized controlled trials. These results therefore suggested that we could pool RCTs and quasi-experimental studies with a medium risk of selection-bias.

Meta-regressions presented further evidence for the inability to pool randomized controlled trials and quasi-experimental studies with a high risk of selection-bias. The estimated effect sizes on economic empowerment of RCTs were substantively and statistically significantly lower than the effect sizes of quasi-experimental studies with a high risk of selection-bias ($\beta=-0.44$; 95% CI=-0.81, -0.07). At the same time, meta-regression indicated that the estimated effect sizes on economic empowerment of RCTs were not statistically significantly different from the effect sizes of quasi-experimental studies with a medium risk of selection-bias ($\beta=-0.04$; 95% CI=-0.09, 0.29). Based on these meta-analyses and meta-regressions we decided to only pool randomized controlled trials and quasi-experimental evaluations with a medium risk of selection-bias.

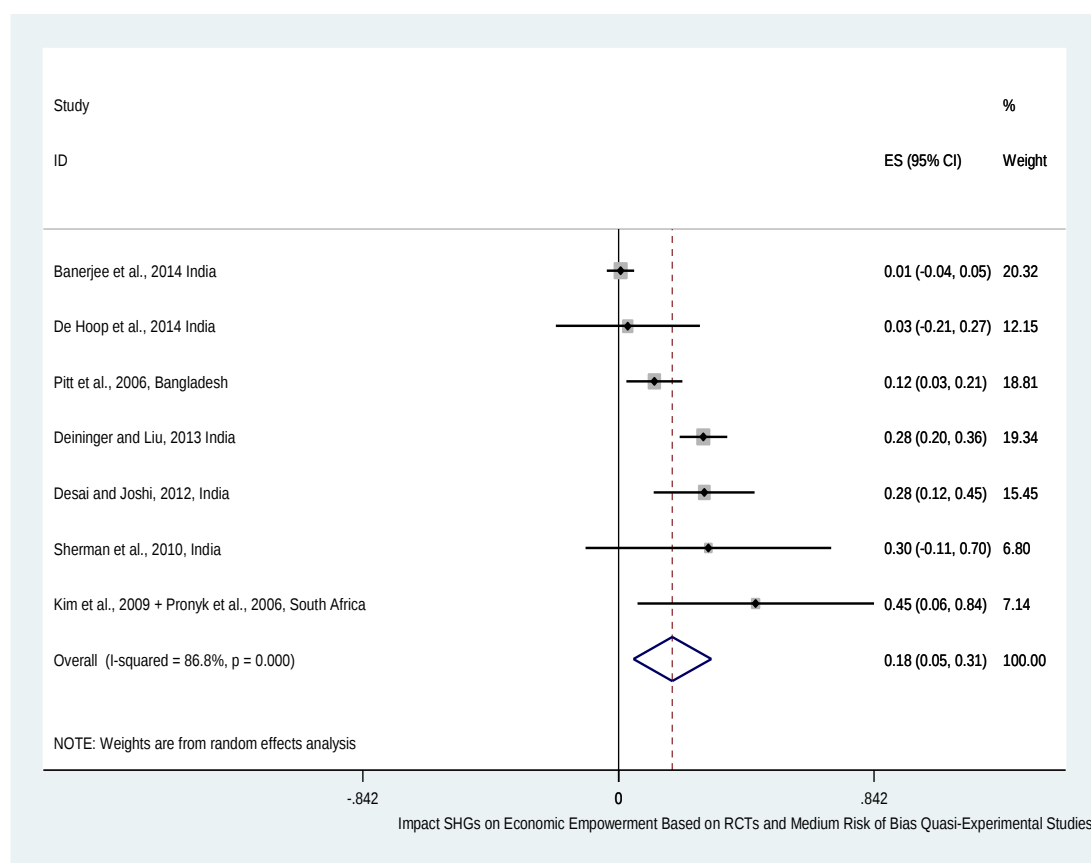
Further analyses of the quasi-experimental studies with a medium risk of selection-bias did not suggest evidence for differences in estimated effect sizes between evaluated self-help groups with and without a training component. A meta-regression indicated that the estimated effect sizes on economic empowerment of quasi-experimental studies with a medium risk of selection-bias for SHGs with a training component are 0.06 SD higher than quasi-experimental studies with a medium risk of selection-bias focusing on SHGs without a training component. The results were, however, not statistically distinguishable from each other at the 5 per cent significance level ($\beta=0.06$; 95% CI=-0.33, 0.45).

The meta-analysis of quasi-experimental evaluations with a medium risk of selection-bias also indicated that studies with a high risk of spillovers might underestimate the impact of women's self-help groups on women's economic empowerment possibly because of contamination of the comparison group. A meta-regression indicated that the estimated effect size of quasi-experimental studies with a medium risk of selection-bias and a high risk of performance bias is statistically and significantly lower than the estimated effect size of quasi-experimental studies with a medium risk of selection-bias and a low or medium risk of performance bias ($\beta=-0.17$; 95% CI=-0.06, -0.28). We explore this relationship further in the pooled analysis of randomized controlled trials and quasi-experimental studies with a medium risk of selection-bias.

Finally, we conducted meta-analysis of randomized controlled trials and quasi-experimental evaluations with a medium risk of selection-bias to determine the pooled effects of women's self-help groups on women's economic empowerment (Figure 4.6). The analysis suggests that women's self-help groups have a positive effect of 0.18 standard deviations on women's

economic empowerment. The effect is statistically significant at the 95 per cent level (SMD=0.18, 95% CI=0.05, 0.31; evidence from 7 studies). The analysis also indicated strong statistical heterogeneity in the impact estimates ($Q=46$, $\text{Tau-sq}=0.02$, $I\text{-sq}=87\%$) with effect sizes ranging from 0.01 to 0.45 standard deviations.

Figure 4. 6: Effects of women’s self-help groups on women’s economic empowerment (RCTs and quasi-experimental evaluations with a medium risk of selection-bias)



Additional analyses suggested that the heterogeneity in the effect sizes were partly explained by training. Analysis of the effects of women’s self-help groups on women’s economic empowerment excluding interventions without a training component (Figure 11.3 in Appendix 11), suggests groups with a training component have a statistically significant positive effect of 0.26 standard deviations on women’s economic empowerment (SMD=0.26, 95% CI=0.17, 0.35; $Q=5$, $\text{Tau-sq}=0.00$, $I\text{-sq}=17\%$; evidence from 5 studies). Furthermore, meta-regression suggested that the effect size of studies with a low or medium risk of selection bias that focused on interventions with a training component had a statistically significantly larger effect size than studies with a low or medium risk of selection bias that focused on interventions without a training component ($\beta=0.20$; 95% CI=0.06, 0.34). In contrast, the evidence for positive effects on economic empowerment of SHGs without a training component is rather less convincing: the average effect size estimated was only 0.06 SMD and was not statistically significant at the 95 per cent significance level (SMD=0.06, 95% CI=-0.05, 0.16; $Q=5$, $\text{Tau-sq}=0.00$, $I\text{-sq}=78\%$; evidence from 2 studies) (Figure 11.4 in Appendix 11). However, it is hard to interpret this finding, because the quantitative studies provide only very limited information about the contents of the training included in the evaluated SHGs.

Table 4.6 summarizes the results of all meta-analyses with an emphasis on economic empowerment. Interestingly, the study by Sherman et al. (2010), which studied the bargaining power of sex workers towards clients, did not show an effect size that was either substantively or statistically significantly different from the effect sizes of the other studies. The interpretation of our results thus did not change when we excluded this study. Similarly, our results did not change substantively when we excluded the study of Deininger and Liu (2013), which focuses on women's ability to save individually.

We did not find evidence for differences in effect sizes of studies with a low or medium risk of spillovers and studies with a high risk of spillovers in the pooled sample. Our analyses also did not suggest evidence for significant differences in effect sizes between studies with low, medium, and high outcome and analysis reporting and other biases, respectively.

Table 4. 6: Summary of effects of SHGs on economic empowerment

<i>Description</i>	<i>Effect Size</i>	<i>Confidence Interval</i>
Randomized controlled trials	0.22 SMD	-0.01 SMD, 0.44 SMD
Quasi-experimental studies	0.32 SMD	0.14 SMD, 0.50 SMD
RCTs and quasi-experimental studies with medium risk of selection bias	0.18 SMD	0.05 SMD, 0.31 SMD
Quasi-experimental studies with high risk of selection bias	0.65 SMD	0.33 SMD, 0.98 SMD
RCTs and quasi-experimental studies with medium risk of selection bias with an emphasis of SHGs that include training	0.26 SMD	0.17 SMD, 0.35 SMD
RCTs and quasi-experimental studies with medium risk of selection bias with an emphasis of SHGs that do not include training	0.06 SMD	-0.05 SMD, 0.16 SMD

A number of quasi-experimental studies could not be included in the meta-analysis. However, excluding these studies would not have significantly, either substantively or statistically, changed the results from the meta-analysis. Either the results of these studies were not very different from the results of the meta-analysis or the risk of bias of the study would have been too high to be included in the preferred specification for the meta-analysis. Coleman (2002) found positive but small effects of women's self-help groups in Thailand on women's economic empowerment. These results could not be included in the meta-analysis because Coleman (2002) focused on the effects of time in self-help groups rather than the effects of participation in self-help groups. In addition, the study did not focus on women's bargaining power but on women's ownership of assets, such as land, so the outcome indicators were not considered comparable to other studies. Garikipati (2008, 2012) assessed the impact of women's self-help groups on different components of women's empowerment, including women's bargaining power as well as other components of women's economic empowerment, but found no evidence of positive effects. These studies were not included because Garikipati (2008, 2012) used the time in self-help groups rather than the participation in self-help groups as an explanatory variable. Holvoet (2005) found that self-help groups had bigger positive effects on women's economic empowerment when

self-help groups provided training in addition to financial services. However, although the study focused on women's bargaining power, the study was considered high risk of selection bias. Husain et al. (2010) suggested positive effects of women's self-help groups on economic empowerment, including women's bargaining power, but did not present the point estimates regarding the impact of women's self-help groups, and the study was considered high risk of selection-bias. Finally, Mukherjee and Kundu (2012) also suggested a positive effect of women's self-help groups on women's economic empowerment, again including women's bargaining power. However, they did not present the quantitative point estimates, and the study was considered high risk of selection-bias.

4.4.2 Social Empowerment

We also synthesized the effects of women's self-help groups on women's social empowerment using meta-analysis. Of the 23 included quantitative studies, ten included an impact estimate that we were able to include in meta-analysis, and five included an impact estimate for women's social empowerment but did not allow determination of the effect size of the intervention (Table 4.7). Analysis of outcomes indicated that social empowerment relates to two types of outcome variables: 1) outcome variables that are associated with women's mobility; and 2) outcome variables that relate to reproductive behavior and the bargaining power of women over family-size decision-making. We therefore conducted both pooled and stratified meta-analyses of studies according to these constructs.

Table 4. 7: Measurement of women's social empowerment

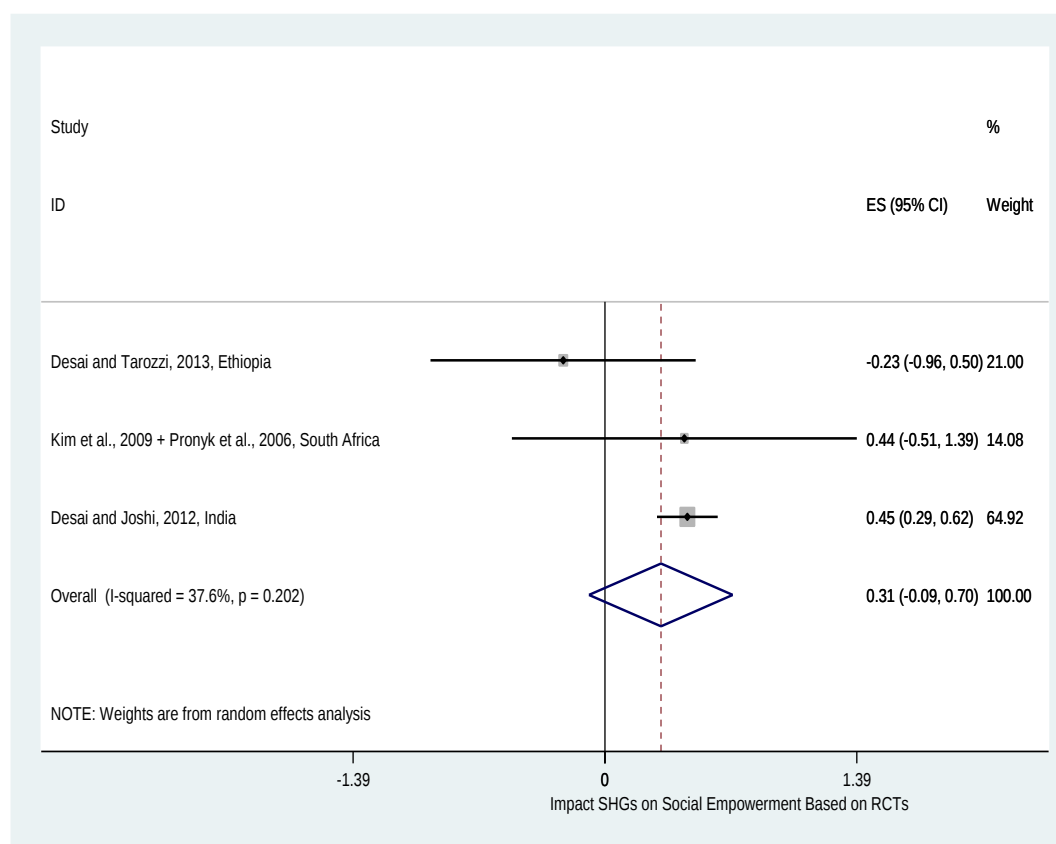
Study	Definition of Variable	Scale	Included in Meta-Analysis?
Bali Swain & Wallentin (2009)	General index of women's empowerment.	Normalized score from 0-1	No; not able to estimate effect size
De Hoop et al., (2014)	Several dummy variables that measure women's autonomy to go out without their husband's permission	Several binary variables	Yes
Deininger and Liu (2013)	Several dummy variables that measure women's autonomy to go out without their husband's permission	Several binary variables	Yes; after imputing the standard deviation
Desai and Tarozzi (2011)	Several dummy variables that measure women's decision-making power about family-size decision-making	Several binary variables	Yes; after calculating pooled effect size
Desai and Joshi (2012)	Dummy variable associated with family-size decision-making	Binary variable	Yes
Husain et al. (2010)	Several variables associated with women's mobility, which are 0 if the woman has no decision-making power, 0.5 if there is joint decision-making and 1 if the woman is the sole decision-maker.	Aggregate score of categorical variables	No; not able to estimate effect size
Kim et al. (2009) & Pronyk et al. (2006)	Dummy variable that is 1 if the woman challenges gender norms	Binary variable	Yes

Study	Definition of Variable	Scale	Included in Meta-Analysis?
Mahmud (1994)	Dummy variable that is 1 if the woman is sterilized	Binary variable	Yes
Nessa et al. (2012)	Categorical variable associated with freedom of movement of woman	Categorical variable	Yes
Pitt et al. (2006)	Several dummy variables that are associated with women's mobility and women's bargaining power over family-size decision-making	Index of binary variables	Yes
Rosenberg et al. (2011)	Several dummy variables that are associated with reproductive behavior and family-size decision-making	Several binary variables	No
Steel et al. (1998)	Dummy variable that is 1 when the woman uses contraceptives	Binary variable	Yes
Swendeman et al. (2009)	Several dummy variables that are associated with reproductive behavior	Several binary variables	Yes

Our meta-analysis commenced with the synthesis of results from randomized controlled trials. The meta-analysis was based on three studies, two of which showed close to identical point estimates. However, the analysis also indicated strong heterogeneity in the impact estimates (Figure 4.7). The effect sizes ranged from -0.23 to 0.45 standard deviations, and the pooled effect size was not statistically significantly different from zero (SMD=0.31, 95% CI=-0.09, 0.70; Q=3, Tau-sq=0.06, I-sq=38%; evidence from 3 studies).

There were also potentially important differences between the three studies included in the meta-analysis. Two of the studies focused on family-size decision-making (Desai & Tarozzi, 2011; Desai & Joshi, 2012), while the study of Kim et al. (2009) presents the impact of a self-help group on an outcome variable associated with the challenging of gender norms by the women respondents. Unfortunately, the latter outcome variable was not very well explained in the paper, but we interpret it as being associated with women's family-size decision-making because the intervention mostly focused on that aspect of women's social empowerment. Second, each of the studies took place in a different part of the world. The study of Desai and Tarozzi (2011) focused on Ethiopia, while the study of Kim et al. (2009) presented impact estimates in the setting of South Africa. Finally, Desai and Joshi (2012) focused on the impact of women's self-help groups on women's social empowerment in the context of India. Third, although the studies of Desai and Joshi (2012) and Kim et al. (2009) both include a training component, the study of Desai and Tarozzi (2012) focused on a self-help group intervention without a training component. Fourth, there were differences in the risk of bias assessment across the three studies. Clearly, the sheer number of differences between the three different studies made it impossible to explain the differences in the effect sizes across the three studies based on a quantitative analysis alone. Therefore, we refrained from undertaking a meta-regression to examine the differences in the effect sizes.

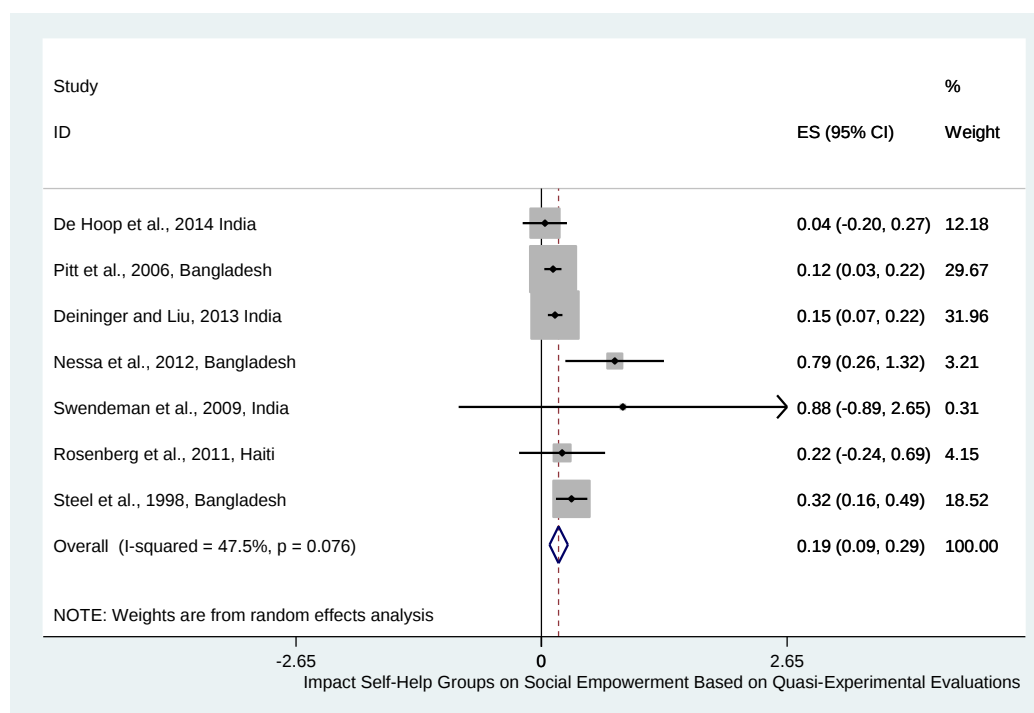
Figure 4. 7: The effect of women's self-help groups on women's social empowerment (randomized controlled trials)



We interpreted the findings of the RCTs as evidence for positive effects of SHGs on women's family-size decision-making power and not of evidence for positive effects on women's mobility. None of the RCTs focused on women's mobility. In later stages of our analysis we found evidence that women's family-size decision-making power and women's mobility should not be considered part of the same construct.

We also conducted meta-analysis of quasi-experimental evaluations examining the effects of women's self-help groups on women's social empowerment (Figure 4.8). The analysis suggested that self-help groups have a positive effect on women's social empowerment. The point estimate of 0.19 standard deviations is significant at the 95 per cent significance level (SMD=0.19, 95% CI=0.09, 0.29; evidence from 7 studies). The results also indicated significant statistical heterogeneity ($Q=3$, $\text{Tau-sq}=0.01$, $I\text{-sq}=48\%$) and the effect size ranged from 0.04 to 0.88 standard deviations across studies.

Figure 4. 8: Effects of women's self-help groups on women's social empowerment (quasi-experimental evaluations)

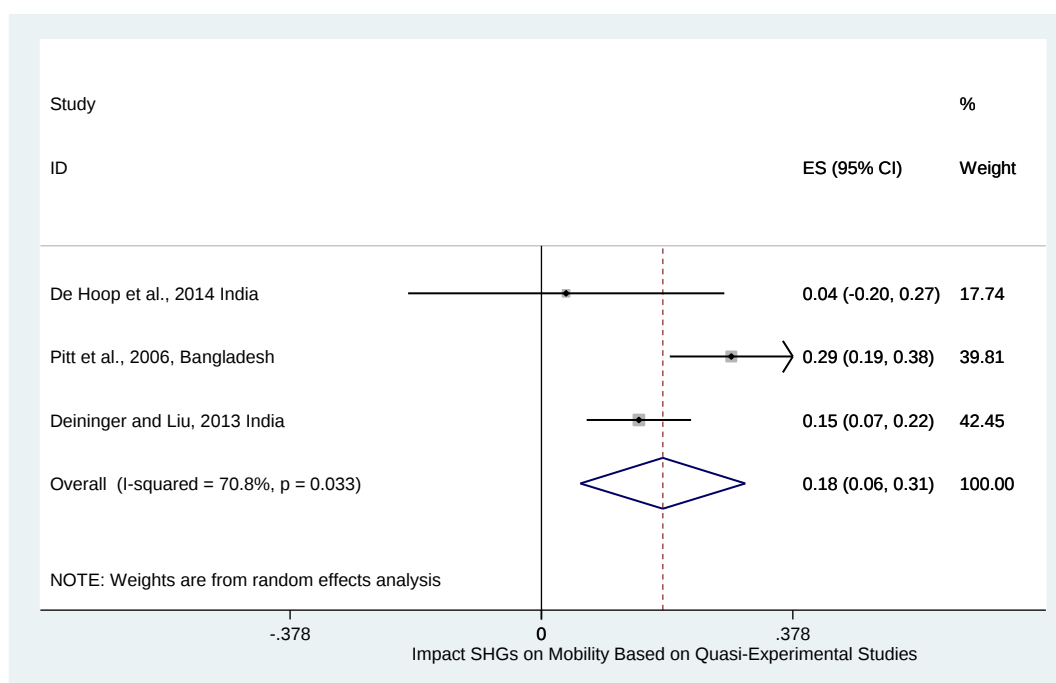


Our analyses suggested that part of the heterogeneity in the effect sizes can be explained by differences in the risk of selection-bias across quasi-experimental studies. We found strong differences between the effect sizes of studies with a high and medium risk of selection-bias, respectively (Figure 11.5 and 11.6 in Appendix 11). The average effect size for quasi-experimental studies with a high risk of selection-bias was estimated at 0.37 standard deviations (SMD=0.37, 95% CI=0.18, 0.56; Q=3, Tau-sq=0.01, I-sq=10%; evidence from 4 studies), and statistically significantly different from the average of 0.13 standard deviations for quasi-experimental studies with a medium risk of selection bias (SMD=0.13, 95% CI=0.07, 0.19; Q=1, Tau-sq=0.00, I-sq=0%; evidence from 3 studies). Our analysis thus suggested studies with a high risk of selection-bias were biased and should not be pooled with studies with a medium risk of selection-bias. Meta-regression confirmed that the estimated effect size for quasi-experimental studies with a high risk of selection-bias was significantly higher than the estimated effect size on social empowerment of quasi-experimental studies with a medium risk of selection-bias ($\beta=0.22$, 95% CI=0.06, 0.39). Based on these analyses we concluded that, while we could pool quasi-experimental studies with a medium risk of selection-bias with randomized controlled trials, we should not pool these studies alongside quasi-experimental studies with a high risk of selection-bias.

We also estimated stratified meta-analyses for quasi-experimental studies focusing on women's family-size decision-making and women's mobility, respectively. We found large and positive pooled effects of studies with a high risk of selection bias on women's family-size decision-making (SMD=0.53, 95% CI=0.22, 0.85; Q=18, Tau-sq=0.08, I-sq=83%; evidence from 4 studies) (Figure 11.7 in Appendix 11). However, the results are likely to be biased because the estimates are significantly larger than the impact estimates of Pitt et al. (2006), the only quasi-experimental study with a medium risk of selection bias that focuses on family-size decision-making (SMD=0.06, 95% CI=-0.04, 0.15). Analysis of studies of effects on women's mobility, of which only quasi-experimental studies with a medium risk of

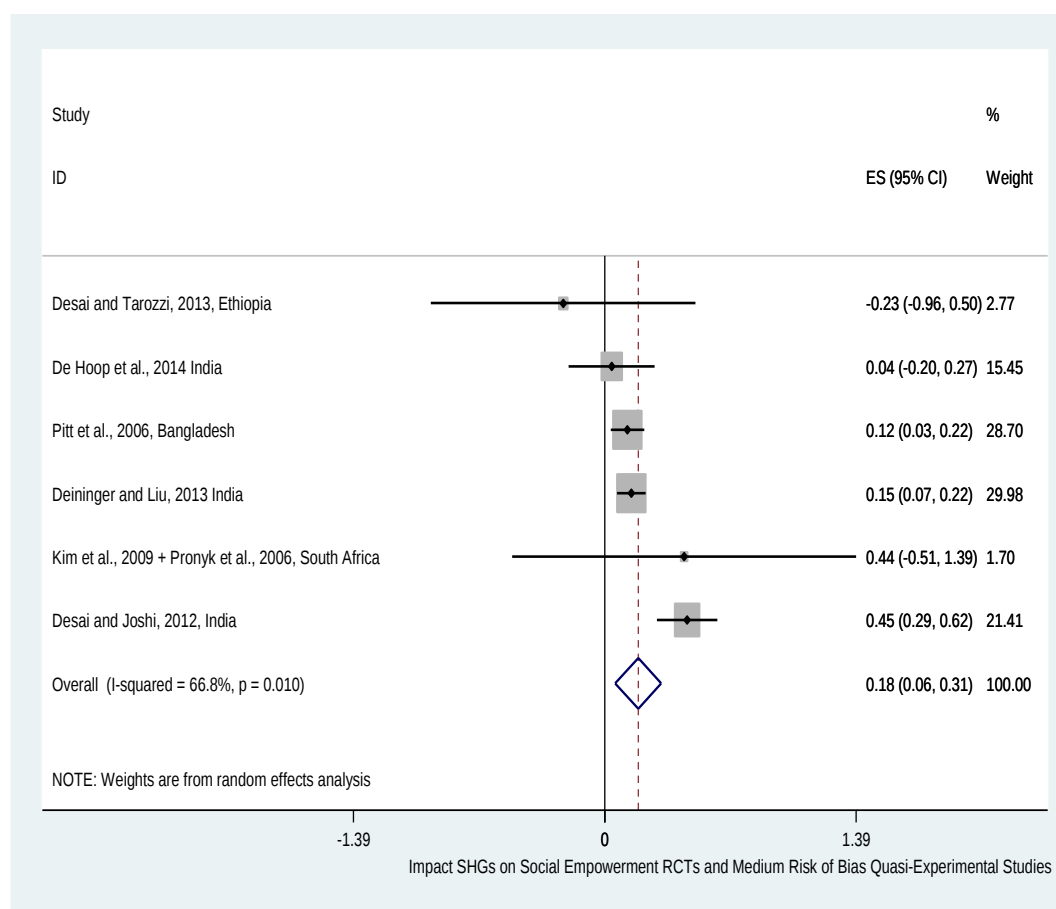
selection bias were available, suggested positive and statistically significant effects (SMD=0.18, 95% CI=0.06, 0.31; Q=7, Tau-sq=0.01, I-sq=71%; evidence from 3 studies) (Figure 4.9).

Figure 4. 9: Effects of women's self-help groups on women's mobility (quasi-experimental evaluations)



Finally, we estimated the pooled effects of self-help groups on social empowerment across randomized controlled trials and quasi-experimental studies with a medium risk of selection-bias. The analysis suggested an average positive and statistically significant effect of 0.18 standard deviations (SMD=0.18, 95% CI=0.06, 0.31; Q=15, Tau-sq=0.01, I-sq=67%; evidence from 6 studies) (Figure 4.10).

Figure 4. 10: Effects of women's self-help groups on women's social empowerment (RCTs and quasi-experimental evaluations with a medium risk of selection-bias)



Interestingly, all RCTs included in the meta-analysis focused on women's bargaining power over family-size decision-making. Almost all quasi-experimental studies focused only on women's mobility. Only the study of Pitt et al. (2006) presented a weighted average estimate for women's social mobility and family-size decision-making. The difference in emphasis between RCTs and quasi-experimental studies might explain why randomized controlled trials tend to show a larger effect on social empowerment than quasi-experimental studies with a medium risk of selection-bias.

Meta-analysis results also indicated that SHGs have a stronger effect on women's family-size decision-making power than on women's mobility (Figures 4.11 and 4.12). The average effect of SHGs on women's family-size decision-making power appears to be 0.26 standard deviations (SMD=0.26, 95% CI=-0.04, 0.56; Q=21, Tau-sq=0.07, I-sq=86% ; evidence from 4 studies), while the average effect on women's mobility appears to be 0.18 standard deviations (SMD=0.18, 95% CI=0.06, 0.31; Q=7, Tau-sq=0.01, I-sq=71% ; evidence from 3 studies).

Thus, we interpret this finding as suggesting that SHGs have a larger impact estimate on family-size decision-making than on women's mobility. The larger effect on family-size decision-making was also illustrated by one study which assessed within-study impacts on both women's family-size decision-making power and women's mobility, finding positive effects on women's family-size decision-making but no evidence for positive effects on women's mobility (Pitt et al., 2006).

Figure 4. 11: Effects of women's self-help groups on women's mobility

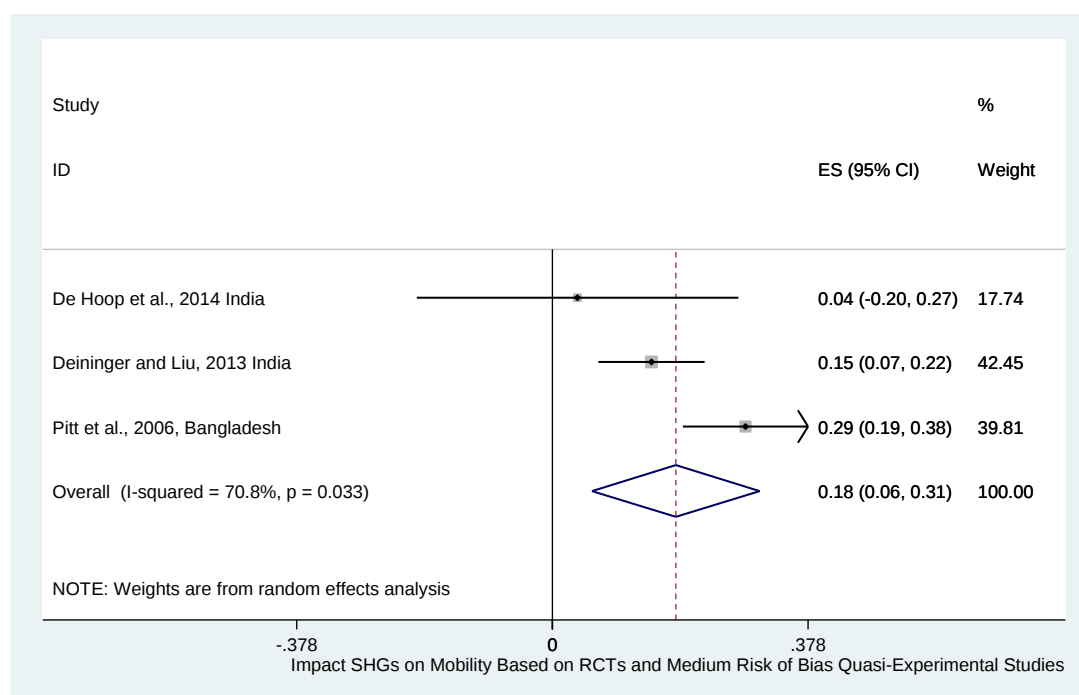
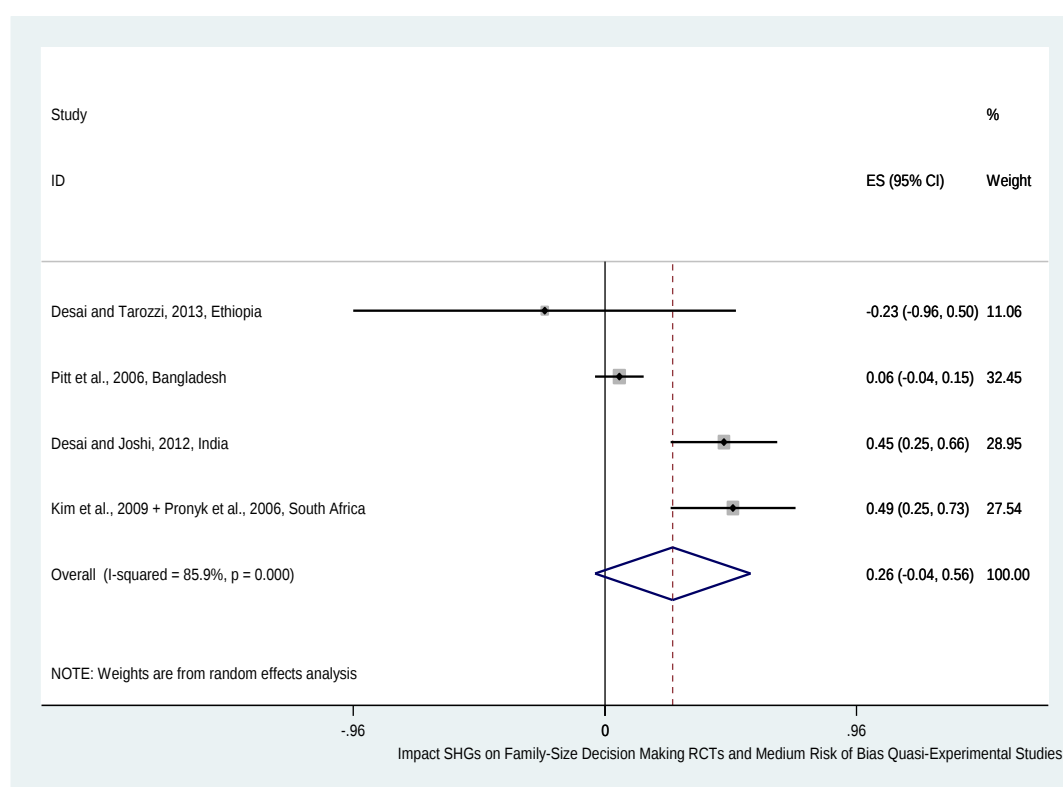


Figure 4. 12: Effects of women's self-help groups on women's family-size decision-making



Our analyses also suggested that training in SHGs might have stronger effects on women's family-size decision-making power than on women's mobility. For family-size decision-making we found that the effect sizes of studies that focus on SHGs with a training element were substantially and statistically significantly higher than the effect sizes of studies without

a training element ($\beta=0.38$; 95% CI=0.19, 0.57). Additional meta-analyses suggested that the effect size of SHGs on family-size decision-making was positive and statistically significant at the 95 per cent significance level when we excluded studies without a training component (SMD=0.41, 95% CI=0.19, 0.63; $Q=3$, $\text{Tau-sq}=0.02$, $I\text{-sq}=41\%$; evidence from 3 studies) (Figure 11.8 in Appendix 11). At the same time the effect sizes remained heterogeneous ranging between -0.23 and 0.49 SMD, suggesting that the type of training was important. Unfortunately, however, the included studies did not present much detail on the type of training. Thus, we have to remain careful in the interpretation of the effects of training in SHGs on women's family-size decision-making power.

For mobility, evidence from meta-regression suggested a counter-intuitive finding, namely that SHGs with a training component had a lower effect on mobility than studies without a training component ($\beta=-0.15$; 95% CI=-0.03, -0.27). However, this finding is driven entirely by a single study – Pitt et al. (2006) is the only study with an emphasis on the effects of SHGs without a training component on women's mobility (SMD=0.29, 95% CI=0.19, 0.38). Furthermore, the findings are counter-intuitive hence we are careful in interpreting them. Figure 11.9 (Appendix 11) presents the meta-analysis for the effect of SHGs on women's mobility for studies with an emphasis on SHGs with a training component. The results show an average effect size of 0.14 SMD that is statistically significantly different from zero at the 95 per cent confidence level (SMD=0.14, 95% CI=0.06, 0.21; $Q=3$, $\text{Tau-sq}=0.00$, $I\text{-sq}=0\%$; evidence from 2 studies).

The findings suggested that women's mobility and women's family-size decision making should not be considered as part of the same construct. Thus, in summarizing the findings the effects of SHGs on social empowerment, we separate women's mobility and family-size decision-making (Tables 4.8 and 4.9).

Table 4. 8: Summary effects of SHGs on women's mobility

<i>Description</i>	<i>Effect Size</i>	<i>Confidence Interval</i>
Randomized controlled trials	N/A	N/A
Quasi-experimental studies	0.18 SMD	0.06 SMD; 0.31 SMD
RCTs and quasi-experimental studies with medium risk of selection bias	0.18 SMD	0.06 SMD; 0.31 SMD
Quasi-experimental studies with high risk of selection bias	N/A	N/A
RCTs and quasi-experimental studies with medium risk of selection bias with an emphasis on SHGs that included training	0.14 SMD	0.06 SMD; 0.21 SMD
RCTs and quasi-experimental studies with medium risk of selection bias with an emphasis on SHGs that did not include training	0.29 SMD	0.19 SMD; 0.38 SMD

Table 4. 9: Summary effects of SHGs on women's family-size decision-making power

<i>Description</i>	<i>Effect Size</i>	<i>Confidence Interval</i>
Randomized controlled trials	0.31 SMD	-0.09 SMD; 0.70 SMD
Quasi-experimental studies	0.06 SMD	-0.04 SMD;0.15 SMD
RCTs and quasi-experimental studies with medium risk of selection bias	0.25 SMD	-0.03 SMD;0.54 SMD
Quasi-experimental studies with high risk of selection bias	0.53 SMD	0.22 SMD;0.85 SMD
RCTs and quasi-experimental studies with medium risk of selection bias with an emphasis on SHGs that include training	0.41 SMD	0.19 SMD;0.63 SMD
RCTs and quasi-experimental studies with medium risk of selection bias with an emphasis on SHGs that do not include training	0.06 SMD	-0.04 SMD;0.15 SMD

The included studies with an emphasis on social empowerment only included one study with an emphasis on social empowerment that was not included in the meta-analysis. This paper did not report an effect size but found positive effects on women's mobility (Husain et al., 2010), consistent with the meta-analysis.

Furthermore, the distribution of effect sizes in the meta-analysis gives some indication for a relationship between contextual characteristics and the impact of women's self-help groups on women's family-size decision-making power. We analyze this distribution of effect sizes more carefully using a narrative analysis because our sample size did not allow for a stratified meta-analysis or meta-regression. The study in Ethiopia showed the least convincing evidence for positive effects on women's family-size decision making power (Desai & Tarozzi, 2014). At the same time the self-help group in Rajasthan, India, showed strong effects on women's family-size decision-making (Desai & Joshi, 2012). The results suggest there may be a difference in the effects of self-help groups on women's social empowerment across regions. We will further explore this mechanism in the qualitative analysis.

4.4.3 Political Empowerment

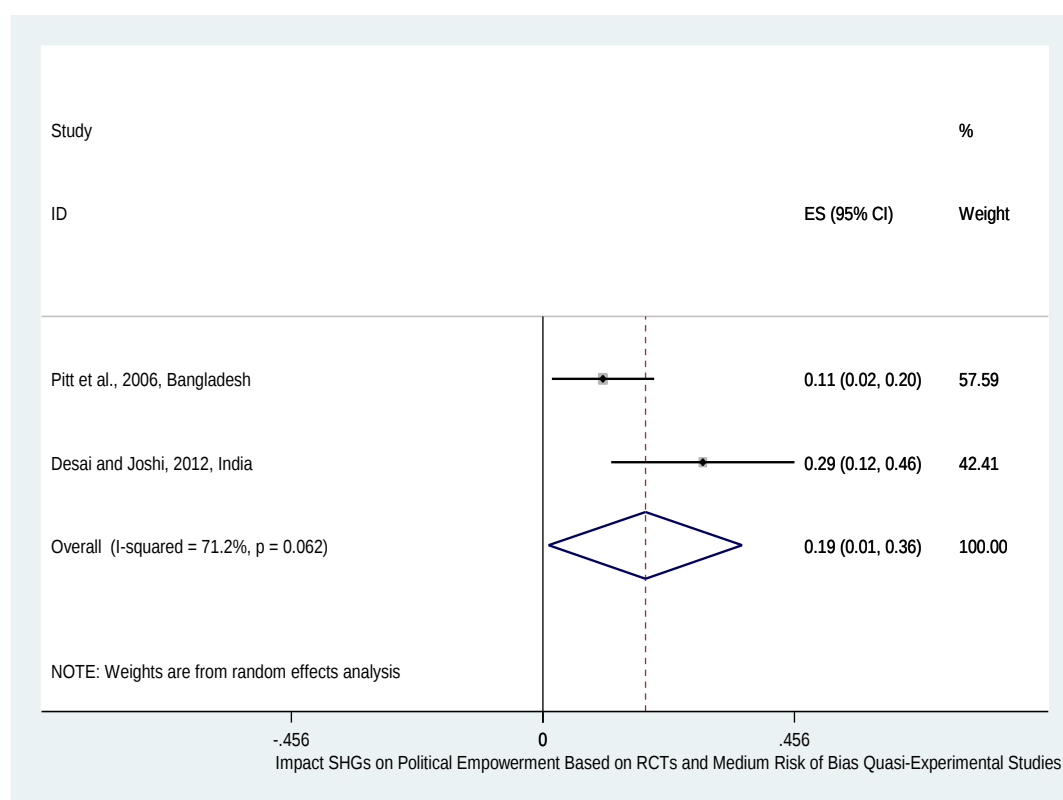
We were able to include 23 quantitative studies which estimated the effects of women's self-help groups on women's political empowerment. Of these, three included an estimate of women's political empowerment resulting from SHGs that we were able to include in our meta-analysis. However, we only included two effect sizes because including the study of Swendeman et al. (2009) may result in bias due to the high risk of selection-bias. Furthermore, we were unable to determine the effect size for one study estimating the impact of SHGs on women's political empowerment (table 4.10).

Table 4. 10: Measurement of political empowerment

Study	Definition of Variable	Scale	Included in Meta-Analysis?
Deininger and Liu (2013)	Several dummy variables that are associated with women's voting behavior	Several binary variables	No; not able to estimate effect size
Desai and Joshi (2012)	Several dummy variables that are associated with women's participation in community meetings and elections	Several binary variables	Yes
Pitt et al. (2006)	Several dummy variables that are associated with women's voting behavior	Index of binary variables	Yes
Swendeman et al. (2009)	Several dummy variables that are associated with women's voting behavior	Several binary variables	No, high risk of selection-bias

For our meta-analysis to determine the effects of women's self-help groups on political empowerment, we decided to pool one RCT and the quasi-experimental evaluation with a medium risk of selection bias for which we were able to estimate the effect size in one meta-analysis (Pitt et al., 2006; Desai and Joshi, 2012). We pooled these studies because our previous analyses to determine the effects of SHGs on economic and social empowerment suggest that studies with a low-or medium risk of selection bias can be pooled in one meta-analysis without biasing the results. We did not include the study of Swendeman et al. (2009) with a high risk of selection-bias in our meta-analysis because the evidence from the meta-analysis on economic and social empowerment indicated that studies with a high risk of selection-bias have an upward bias. Although we were not able to gain a nuanced understanding of the impacts of women's self-help groups based on the two studies that we were able to include in meta-analysis, the results suggested that women's self-help groups have a positive effect on women's political empowerment. The average effect of women's self-help groups on political empowerment was estimated as 0.19 standard deviations (SMD=0.19, 95% CI=0.01, 0.36; Q=3, Tau-sq=0.01, I-sq=71%; evidence from 2 studies) (Figure 4.13). The limited number of studies did not allow for sensitivity analysis.

Figure 4. 13: Effects of women's self-help groups on political empowerment



The study of Swendeman et al. (2009) also finds positive effects of SHGs on women's political empowerment. Although this study was not included in our meta-analysis because of the high risk of selection-bias, the positive effect on women's political empowerment is consistent with the findings from our meta-analysis.

The study of Deininger and Liu (2009) also included an estimate on political empowerment. However, it remained unclear how political empowerment was defined in that version of the paper. Furthermore, the published paper (Deininger & Liu, 2013) did not include a focus on political empowerment. It appeared as if the political empowerment variable from the working paper included elements of social empowerment. Therefore, the results of the paper, which reported positive effects on political empowerment, were not included in our meta-analysis. Nonetheless, the positive effects are consistent with the findings of the meta-analysis.

4.4.4 Psychological Empowerment

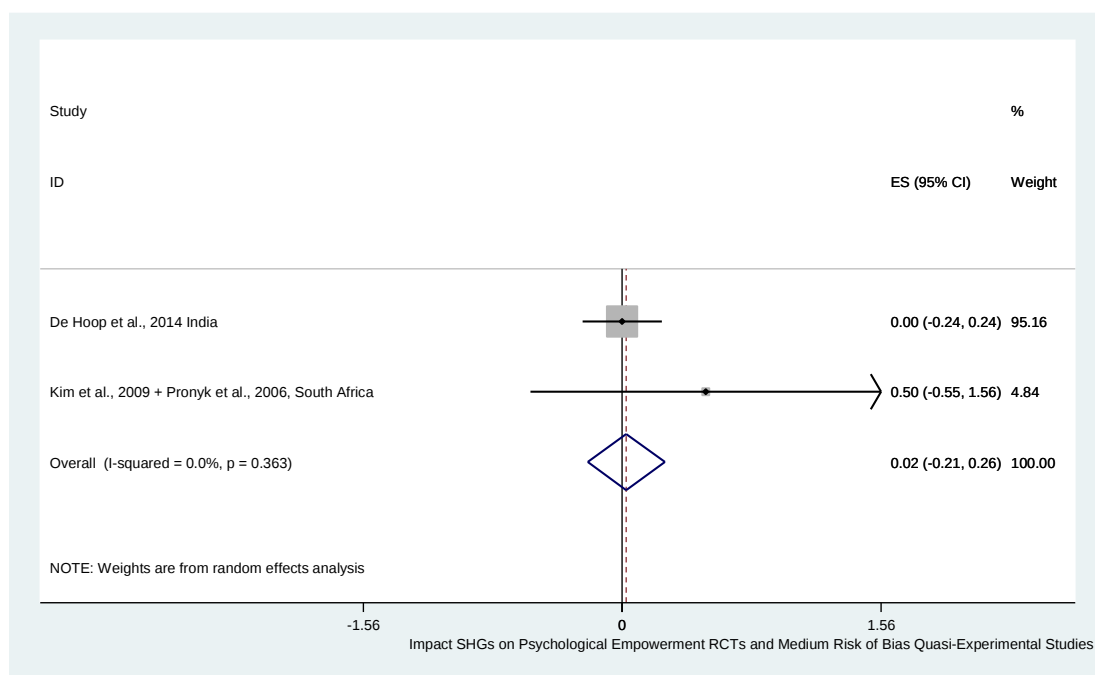
Finally, we synthesized the effects of women's self-help groups on women's psychological empowerment across studies, including adverse effects. Of the 23 included quantitative studies, three included an impact estimate on women's psychological empowerment that we were able to include in our meta-analysis. However, we only included two studies in our meta-analysis because including the study by Swendeman et al. (2009) may result in bias due to the high risk of selection-bias. (Table 4.11).

Table 4. 11: Measurement of psychological empowerment

Study	Definition of Variable	Scale	Included in Meta-Analysis?
De Hoop et al. (2014)	Five-point Likert scale ranging from highly disagree to highly agree as a response to the statement "I have control over my own life"	Categorical variable	Yes
Kim et al. (2009) & Pronyk et al. (2006)	Dummy variable that is 1 if the respondent reports to be self-confident	Binary variable	Yes
Swendeman et al. (2009)	Dummy variable that is 1 when the sex worker reports that sex work is valid work	Binary Variable	No, high risk of selection-bias

As in the meta-analysis for political empowerment, we pooled RCTs and quasi-experimental studies with a medium risk of selection-bias in our meta-analysis for psychological empowerment (De Hoop et al., 2014; Kim et al., 2009), but we did not include studies with a high risk of selection-bias (Swendeman et al., 2009). Although our meta-analysis is only based on two studies, the forest plot in Figure 4.14 indicated that there is major heterogeneity in the effects of women's self-help groups on psychological empowerment (SMD=0.02, 95% CI=-0.21, 0.26; Q=1, Tau-sq=0.00, I-sq=0%; evidence from 2 studies). One study in India did not find positive effects on psychological empowerment (De Hoop et al., 2014). A second study in South Africa demonstrated a large point estimate, but the sample size was too small and consequently the confidence interval too wide to derive strong conclusions regarding the effect of women's self-help groups on psychological empowerment (Kim et al., 2009). Arguably, there is no evidence for positive effects of SHGs on psychological empowerment based on the studies we included in our meta-analysis.

Figure 4. 14: Effects of women's self-help groups on psychological empowerment



4.4.5 Intimate Partner Violence and Other Potential Adverse Effects

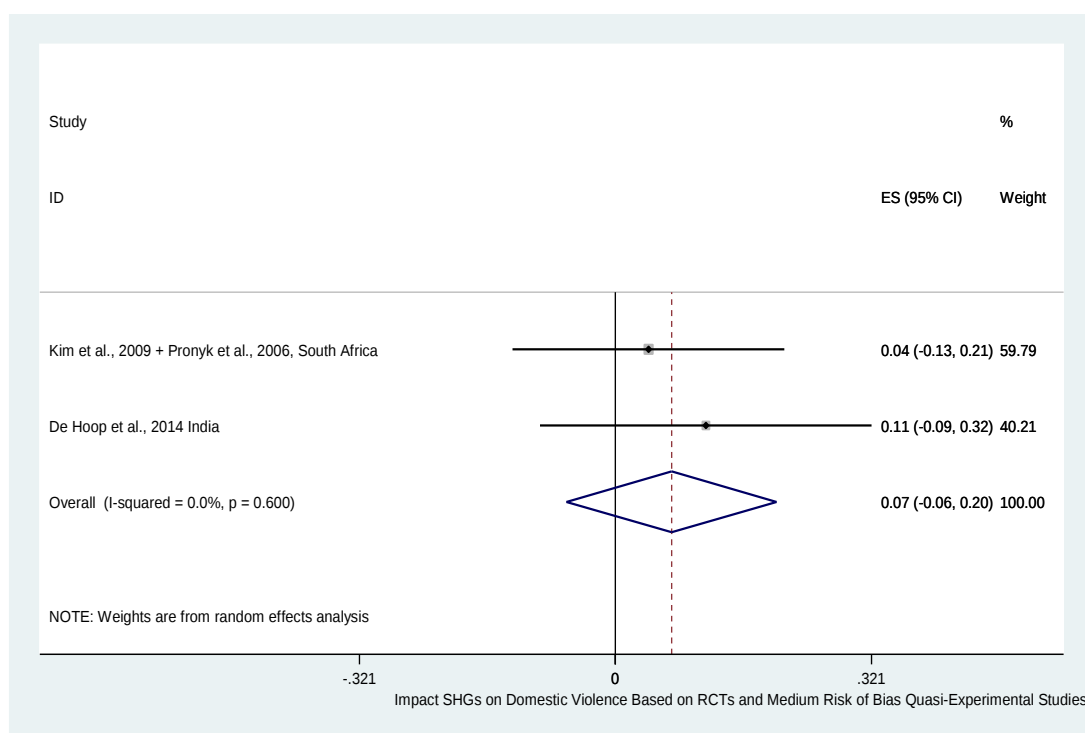
We also synthesized the adverse effects of women's self-help groups with a strong focus on intimate partner violence. Of the 23 included quantitative studies, three included an impact estimate on intimate partner violence that we were able to include in our meta-analysis, and one estimated the impact on partner violence, but we were not able to determine the effect size of this study (Table 4.12). However, as in our previous meta-analyses with few studies we do not include studies with a high risk of selection-bias in the meta-analysis (Ahmed, 2005).

Table 4. 12: Measurement of intimate partner violence

Study	Definition of Variable	Scale	Included in Meta-Analysis?
Ahmed (2005)	Dummy variable that is 1 if the respondent reports that she has been a victim of any type of violence	Binary	No, high risk of selection-bias
De Hoop et al., (2014)	Five-point Likert scale ranging from highly disagree to highly agree as a response to the statement "Men are entitled to beat their women in certain occasions"	Categorical variable	Yes
Husain et al. (2010)	Several variables associated with women's tolerance of domestic violence, which are 0 if the woman thinks violence is not justified, 0.5 if the woman is uncertain and 1 if the woman thinks violence is justified	Several categorical variables	No; not able to estimate effect size
Kim et al. (2009) & Pronyk et al. (2006)	Several dummy variable that are 1 if the respondent condones intimate partner violence	Several binary variable	Yes

Our theory of change suggests that women's self-help groups might have adverse consequences, in the sense that domestic violence could increase as a result of participation in women's self-help groups. However, the meta-analysis of the effects of women's self-help groups on attitudes toward domestic violence, did not show evidence for adverse effects of women's self-help groups on attitudes toward domestic violence (Figure 4.15). As in our meta-analyses for political and psychological empowerment we only pool RCTs and studies with a medium risk of selection-bias in our meta-analysis (De Hoop et al., 2014; Kim et al., 2009). Although the point estimate suggests a positive effect of SHGs on positive attitudes towards domestic violence the relationship is not statistically significant (SMD=0.07, 95% CI=-0.06, 0.20; Q=0, Tau-sq=0.00, I-sq=0%; evidence from 2 studies). Arguably, our meta-analysis did not allow for a nuanced understanding of the effect of women's self-help groups on intimate partner violence, because we only found two studies with a low or medium risk of selection-bias that could be included in the meta-analysis. More rigorous evidence about the effect of women's self-help groups on intimate partner violence is clearly needed. However, at this moment our meta-analysis does not show evidence for adverse effects of women's self-help groups via a contribution to intimate partner violence.

Figure 4. 15: Effects of women's self-help groups on intimate partner violence



In addition to the studies we included in our meta-analysis, two other studies also presented impact estimates on intimate partner violence. First, Ahmed (2005) presented evidence for an adverse but non-significant effect of women's self-help group membership on the likelihood of female respondents having encountered violence. This study was not included in the meta-analysis because of the high risk of selection-bias. Second, Husain et al. (2010) presented findings that suggested a negative effect of women's self-help group membership on women's tolerance of domestic violence. However, we were not able to estimate the effect size of this study because point estimates were not reported. Furthermore, the study was rated as having a high risk of selection-bias.

Our included studies only contained one study that focused on other adverse consequences of women's self-help groups. De Hoop et al. (2014) argued that, on average, women's self-help groups might not have adverse consequences for subjective well-being or happiness, but, at the same time, they found strong negative effects on happiness of women's self-help group members in relatively conservative areas. De Hoop et al. (2014) argued these negative effects occurred because of social sanctioning of women who show autonomous behavior and because of the internal psychological struggles of women who are autonomous in a patriarchal context where this is not considered appropriate behavior for a woman. The absence of average negative effects in the full sample and the strong negative effects in areas with relatively conservative gender norms indicate that adverse consequences of women's self-help groups may be clouded by heterogeneities in the impact estimates. Alternatively, negative effects may also be underreported or researchers may not focus on collection of data on adverse outcomes. We have to be cautious in interpreting this result, however, because the finding was based on only one study with a medium risk of selection bias and a high risk of spillovers. Thus, the findings of the study might not be internally or externally valid, although they were supported by qualitative accounts of women's empowerment trajectories reported in the same study.

Table 4.13 summarizes the results of all meta-analyses with an emphasis on political and psychological empowerment or intimate partner violence.

Table 4. 13: Summary effects of SHGs on women’s political and psychological empowerment and intimate partner violence

<i>Description</i>	<i>Effect Size</i>	<i>Confidence Interval</i>
RCTs and quasi-experimental studies with medium risk of selection bias focusing on political empowerment	0.19 SMD	0.01 SMD; 0.36 SMD
RCTs and quasi-experimental studies with medium risk of selection bias with focusing on psychological empowerment	0.02 SMD	-0.21 SMD; 0.26 SMD
RCTs and quasi-experimental studies with medium risk of selection bias focusing on intimate partner violence	0.07 SMD	-0.06 SMD; 0.20 SMD

4.3.5 Publication bias assessment

We relied on funnel plots to determine the potential for publication bias of studies that focused on economic and social empowerment. As discussed above, the number of studies that focused on political and psychological empowerment was not sufficient to determine the potential for publication bias of studies that focused on these topics. For social empowerment we decided not to test for publication bias for women’s family-size decision making power and mobility separately, despite the fact that our meta-analysis suggests that these two empowerment components can be considered different constructs, because this would have resulted in a strong reduction of statistical power to reject the null hypothesis of no publication bias.

Figure 4.16 presents a funnel plot for studies that focused on economic empowerment with a low or medium risk of selection-bias. The basic idea of a funnel plot is that publication bias is most likely when the effect sizes of studies do not follow a normal distribution. As can be clearly seen in the figure, the effects on economic empowerment are not normally distributed. Instead, it appears as if the results are skewed to the right. Hence, the funnel plot suggests that there might be publication bias in the studies that estimated impacts on economic empowerment. For social empowerment we find a similar pattern with results skewed to the right. Thus, there may also be publication bias for impact evaluations that focus on the effects of women’s self-help groups on social empowerment. Funnel plots can be interpreted in multiple ways, however, so we should be careful in interpreting the figure. We can only say there is potential for publication bias in the impact estimates on economic empowerment.

Figure 4. 16: Funnel plot of economic empowerment outcome

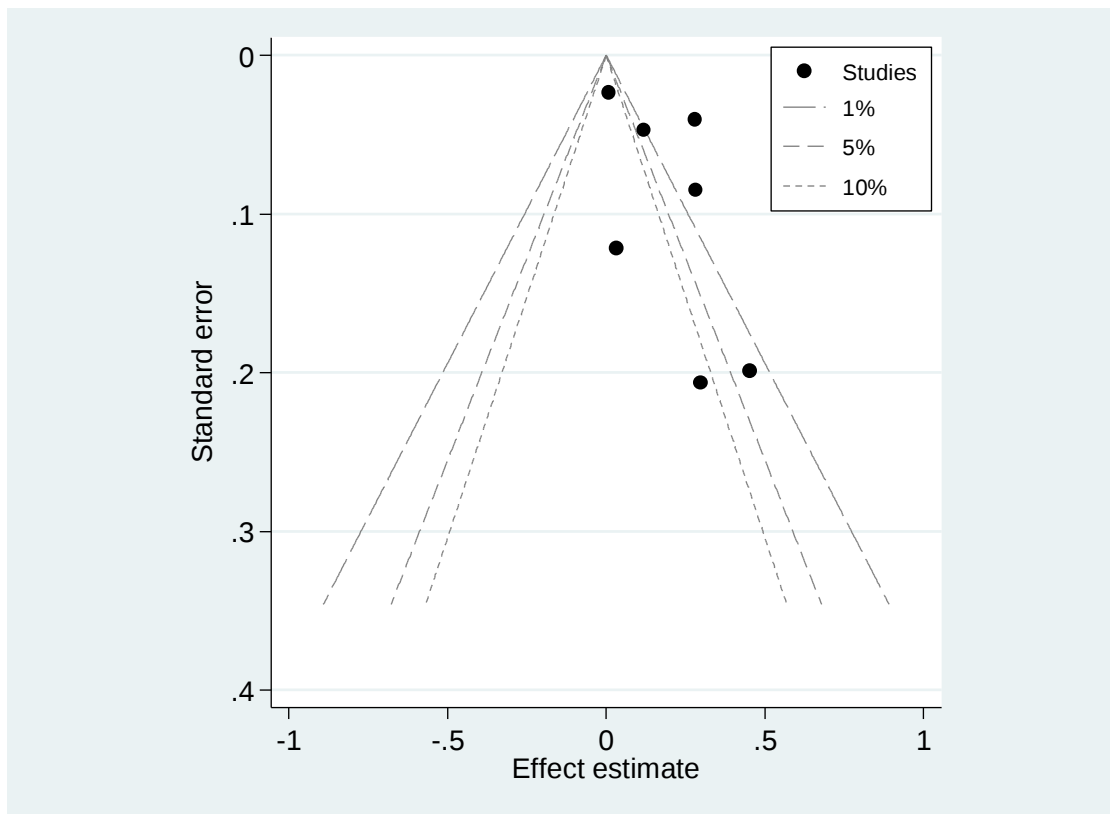
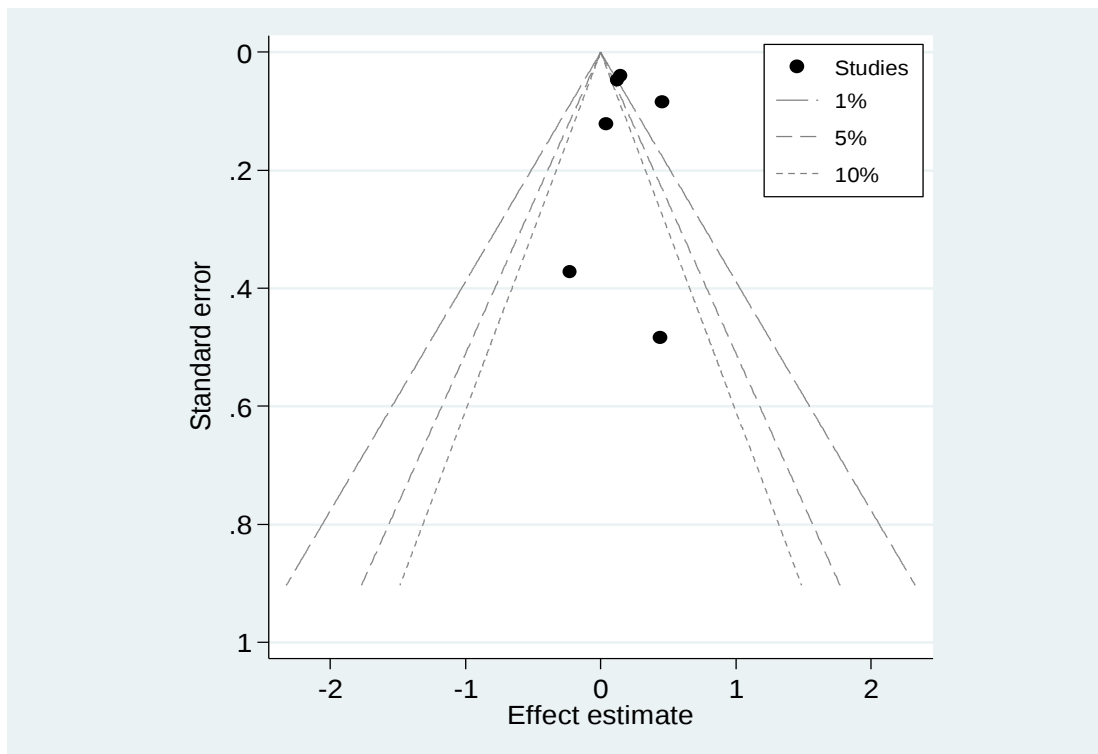


Figure 4. 17: Funnel plot of social empowerment outcome



We formally tested for the potential of publication using Egger's test. For both economic and empowerment we found no formal evidence for publication bias based on this test. For economic empowerment the point estimate for publication bias is positive but the results are not statistically significant ($\beta=2.32$, S.E.=1.58, $p=0.20$). For social empowerment the Egger test indicated no evidence for publication bias ($\beta=0.25$, S.E.=1.33; $p=0.86$). Hence, although there are indications of publication bias in the studies that focused on economic empowerment we found no formal evidence for publication bias based on the Egger test.

Nevertheless, our risk of bias assessment did present some evidence for publication bias. For example, we found two studies that did not report point estimates because the results were not statistically significant (Mahmud, 1994; Steele et al., 1998). This indication of outcome and analysis reporting biases may indicate that the positive impacts we found could be slightly overestimated. Similarly, only a few of our studies (Ahmed, 2005; De Hoop et al., 2014; Husain et al., 2010; Kim et al., 2009) assessed adverse consequences of SHGs. The relatively low number of studies focusing on adverse consequences may indicate reporting bias. However, despite the potential for outcome and analysis reporting bias, we did not find evidence for differential effects for studies with high outcome and analysis reporting bias.

4.5. Synthesis of qualitative studies

The meta-ethnographic analysis of the qualitative studies focused on women's explanations of empowerment outcomes (review objective 2). The 11 studies included in the qualitative analysis came from SHGs in South Asia (Bangladesh, India and Nepal), Bolivia and Tanzania. Table 4.14 and Table 4.15 summarize the findings from the qualitative studies after relying on the meta-ethnographic approach. The following descriptions of the four major outcome categories (economic, social, political, and psychological empowerment) emerged from women's accounts of their self-help group experiences from the 11 contributing studies. A table of additional quotes for each theme is available in Appendix 12.

Table 4. 14: Summary of evidence in qualitative studies

<i>THEME</i>	Dahal 2014, Nepal	Kabeer 2011, Bangladesh	Kilby 2011, South India	Knowles 2014, South India	Kumari 2011, South India	Maclean 2012, Bolivia	Mathrani 2006, South India	Mercer 2002, Tanzania	Pattenden 2011, South India	Ramachandar 2009, South India	Sahu 2012, South India
<i>Psychological Empowerment</i>											
Agentic Voice	x	x	x		x		x			x	
Household Negotiations	x	x			x			x		x	
Impact on Domestic Disputes	x	x	x		x		x			x	x
<i>Social Empowerment</i>											
Improved Networking		x	x	x	x		x		x		x
Solidarity	x	x			x		x				
Community Respect	x	x			x					x	x
<i>Economic Empowerment</i>											
Financial Skills	x	x			x	x		x		x	x
Financial Experience				x	x	x	x	x	x	x	
<i>Political Empowerment</i>											
Broader Social Action		x	x	x	x		x		x	x	x
Limits of Political Context							x		x	x	
<i>Adverse Outcomes</i>											
Barriers to Participation	x						x	x			
Disappointment					x	x	x	x	x		
Corruption	x			x		x		x			
Stigma							x		x	x	

Table 4. 15: Summary of findings from qualitative studies

Theme	Sample Quotes	Contributing Studies	Confidence in Evidence	Explanation of Confidence in Evidence
<u>Psychological Empowerment</u>				
Agentic Voice: Women from South Asia reported feeling more capable of speaking in front of others. Women experienced this by speaking in front of their peers at their group meetings. As groups matured and began to get involved in community development projects, women also talked about feeling capable of speaking in front of others, such as extended families, authorities, and community leaders.	"One of the things I have learned is to be able to speak in front of a group of five people without shivering." Kumari, 2011, South India	Dahal 2014, Kabeer 2011, Kilby 2011, Kumari 2011, Mathrani 2006, Ramachandar 2009	High	Thick data from 6 studies; Only from South Asia; quality was high for 4 studies and medium for 2 studies.
	"My confidence level is increasing. Before, I was afraid to speak out what I disliked, but now I am not dependent on anyone and I can speak my thoughts and I don't care whether someone likes it or not." Dahal, 2014, Nepal			
Participation in Household Decisions: Women discussed the process of gaining acceptance from husbands and in-laws to participate in SHGs. Then, over time, they described gaining respect from husbands and extended family for their contributions and became part of the household decision-making.	"After two years, they [husband and in-laws] understood the value of the women's groups and remained silent." Ramachandar, 2009, South India	Dahal 2014, Kabeer 2011, Kumari 2011, Mercer 2002, Ramachandar 2009, Sahu 2012	High	Thick data from 6 studies; 5 from South Asia, 1 from Tanzania; quality was high for 2 and medium for 4.
	"Being allowed to have money and decide on how to spend it has brought us development in our household and now husbands give us the freedom to do our own things." Mercer, 2002, Tanzania			
Impact on Domestic Disputes: Women reported that their participation had an effect on domestic disputes and violence including both verbal and physical abuse. Women reported an initial increase in disputes or violence but that they eventually gained respect from husbands and in-laws by bringing in income to the household and that they fought less with their husbands. They also reported that their SHGs took action against domestic violence in their communities.	"My husband used to beat me when I became a member of the sangha. He used to manhandle me when I returned home from the meetings. His parents instigated him to beat me. But I stood in silence and today he dare not touch me." Ramachandar, 2009, South India	Dahal 2014, Kabeer 2011, Kilby 2011, Knowles 2014, Kumari 2011, Mathrani 2006, Ramachandar 2009, Sahu 2012	High	Thick data from 8 studies; Only from South Asia; quality was high for 5 and medium for 3.
	"You cannot come drunk and batter me, my SHG will question you if you touch me, you should be prepared to answer them." Kumari, 2011, South India			
<u>Social Empowerment</u>				

Theme	Sample Quotes	Contributing Studies	Confidence in Evidence	Explanation of Confidence in Evidence
Improved Networking: Women SHG members had the confidence to work with local authorities, village leaders, and law enforcers to make positive changes in their communities. These experiences emboldened the women to address authorities when a social issue came up or when they needed support for a community development project. This was a profound change from being confined to the domestic sphere and speaking only to family and close neighbors.	"SHG members complain if a tap is broken or if there is stagnant water ... they bring this to the panchayat [village leader] president's attention issues in the community ... if they have other difficulties they go to government officials now." Knowles, 2014, South India	Kabeer 2011, Kilby 2011, Knowles 2014, Kumari 2011, Mathrani 2006, Pattenden 2011, Sahu 2012	High	Thick data from 7 studies; only from South Asia; quality was high for 4 and medium for 3.
	"The women themselves insisted on dealing with the tractor owners directly and 'held out' for three weeks before the tractor owners agreed to deal with the women directly. It was the close interaction with staff at all levels, which gave the women the confidence to deal with higher caste village people in this way." Kilby, 2011, South India			
Solidarity: Women reported feeling mutual support within their groups and feeling as though they could speak as a collective voice. A sense of solidarity enabled women to make meaningful decisions and to enact positive change in their lives and communities.	"One stick can be broken, a bundle of sticks cannot. It is not possible to achieve anything on one's own. You have no value on your own. Now if I am ill, my [SHG] members will look after me." Kabeer, 2011, Bangladesh	Dahal 2014, Kabeer 2011, Kumari 2011, Mathrani 2006	Moderate	Thick data from 4 studies; only from South Asia; quality was high for 2 and medium for 2.
	"If we disapprove of something, we are able to express our opinions to the larger community as we have a collective voice." Mathrani, 2006, South India			
Community Respect: Women described walking confidently through their villages and feeling respected by their peers and their leaders. They expressed feeling that they were no longer solely housewives but community actors who had influence over their village politics.	"The society's view upon being a SHG member has changed. Before it was against the social norms to go out of a house but now society praises women who are involved in SHGs." Dahal, 2014, Nepal	Dahal 2014, Kabeer 2011, Kumari 2011, Sahu 2012, Ramachandar 2009	High	Thick data from 5 studies, Only from South Asia; quality was high for 2 and medium for 3.
	"The biggest benefit of the [SHG] is that we get prestige and honour in our community; we gain experience going to the bank and meeting with officials." Ramachandar, 2009, South India			
<u>Economic Empowerment</u>				

Theme	Sample Quotes	Contributing Studies	Confidence in Evidence	Explanation of Confidence in Evidence
Financial Skills: Women reported feeling empowered by the newness of handling money. Many of the women had never participated in the buying and selling of goods and had never been allowed to manage the household accounting. With the new access to credit, women were suddenly in the role of the money manager. Women reported that they gained a sense of self-reliance as a result of having access to money, making decisions about buying and selling, and completing transactions with that money	"The fear of handling money is gone." Kumari, 2011, South India	Dahal 2014, Kabeer 2011, Kumari 2011, Maclean 2012, Mercer 2002, Ramachandar 2009, Sahu 2012	High	Thick data from 7 studies across regions (5 from South Asia, 1 from Tanzania and 1 from Bolivia); quality was high for 2 and medium for 5.
	"Being allowed to have money and decide on how to spend it has brought us development in our households and now husbands give us the freedom to do our own things." Mercer, 2002, Tanzania			
Financial Inexperience: Being in charge of finances was a new experience for most women and the women reported feeling unsure about their financial decision making abilities. Some of the SHGs offered training around such topics as income generation and savings. But because women were making decisions in front of their community members, they felt there was a great deal at stake to make sound choices.	"The interest rate is really high. Don Pedro—my husband—tells me off: 'Why are you just working for that [the credit]. You're just working for the bank, and the interest is really expensive!'" Maclean, 2012, Bolivia	Knowles 2014, Kumari 2011, Maclean 2012, Mathrani 2006, Pattenden 2011, Ramachandar 2009	Moderate	Thin description from 6 studies from 2 regions (5 from South Asia and 1 from Bolivia); quality was high for 4 and medium for 2.
	"The men say, 'What kind of structure have these women constructed? They are like monkeys, if we hit their home it will collapse.'" Mathrani, 2006, South India			
<u>Political Empowerment</u>				
Catalyzing Social Action: Women described their participation in a SHG as a "stepping stone" toward wider social participation but not necessarily a political act in itself. Women reported that participation in SHGs did expose them to the concept of women's rights through participation in social activities and it did give them political capital the ability to speak out on political issues such as accountability. Women reported that some members of SHGs went on to become local political leaders.	"In the previous election, the MLA candidate had promised to build a road but he did not. When he came for campaigning this time, we questioned him for not keeping his promise and we didn't vote him either." Sahu, 2012, South India	Dahal 2014, Kabeer 2011, Kilby, 2011 Knowles 2014, Kumari, 2011 Mathrani 2006, Sahu 2012	High	Thick description from 7 studies all from South Asia; quality was high for 4 and medium for 3.
	"SHG members [have] become councillors, government officials ... those elected [in] six out of 15 wards are women and members of elected panchayat bodies. They advanced their skills and were respected by the community." Knowles, 2014, South India			

Theme	Sample Quotes	Contributing Studies	Confidence in Evidence	Explanation of Confidence in Evidence
Understanding the Political Context: SHG members were able to identify the limits to their "empowerment" and described SHGs facing barriers to affecting change in their community through even small political acts. The context within which groups operated "restricted the capacity for political action." Women talked about feeling that awareness of rights was only an important first step and they still had a long way to go before women gained property and reproductive rights. Women agreed that their domestic role of women was still primary.	"Empowerment? There has not been complete empowerment. More factors are needed like equal wages. I would say that only 5 to 10% of empowerment has happened." Kumari, 2011, South India	Kabeer 2011, Kumari 2011, Mathrani 2006, Pattenden 2011, Ramachandar 2009	Moderate	Thin description from 5 studies all from South Asia; quality was high for 3 and medium for 2.
	"Women are still tethered to domesticity and men still regarded women as below them." Kabeer, 2011, South India			
<u>Adverse Outcomes</u>				
Barriers to Participation: Women described barriers to participation specifically for marginalized groups such as lower castes or the very poor. This finding is likely underreported because studies focused on the narratives of participants versus non-participants.	"Some women don't join because they feel inferior, they think that members are rich, can afford things and can be close to the Church, they are in good positions." Mercer, 2002, Tanzania	Dahal 2014, Knowles 2013, Mercer 2002, Mathrani 2006	Moderate	Thin description from 4 studies from two regions (3 from South Asia and 1 from Tanzania); quality was high for 3 and medium for 1.
	"The issue of selection bias can be agreed to a certain extent acknowledging to the fact that very poor people cannot afford the membership fee and enough time for group activities." Dahal, 2014, Nepal			
Disappointment: Women described a degree of disappointment when their groups did not deliver on perceived promises such as solving social problems in their villages like alcoholism. Another source of disappointment occurred when women gained new awareness about rights but were not able to enact them or when their group took on new responsibilities but in the end did not have the authority or financial power to make changes.	"Other women are discouraged because it is almost four to five years since we contributed the money for the cows and up to now we haven't seen any good profit." Mercer, 2002 Tanzania	Dahal 2014, Kabeer 2011, Kumari 2011, Maclean 2012, Mercer 2002	Moderate	Thin description from 5 studies from 3 regions (South Asia, Bolivia and Tanzania); quality was high for 1 and medium for 4.
	"SHGs operate at very low cost, have a small fund, raise little interest so we cannot accomplish bigger projects and this is our weakness." Maclean, 2012, Bolivia			

Theme	Sample Quotes	Contributing Studies	Confidence in Evidence	Explanation of Confidence in Evidence
Mistrust and Corruption: Women reflected on negative experiences such as mistrust and corruption of their group and told stories about corruption they had heard about in other groups specifically of leaders stealing group funds	"I don't like [to be treasurer]. It's dangerous. The money can disappear, you can get confused. Even Dona Feliza [a younger woman who was educated in la Paz] can get a little confused sometimes. And they talk about the treasurer and accuse her of things." Maclean, 2012, Bolivia	Knowles 2014, Maclean 2012, Dahal 2014	Low	Thin description from 3 studies from two regions (South Asia and Bolivia); quality was high for 2 and medium for 1.
	"Accounts are not maintained. The leaders of SHGs are heard to have lent the saved amounts to others at high interest rates for personal benefit." Dahal, 2014, Nepal			
Stigma: Membership in SHGs had negative associations and women faced public shame or discrimination, especially during the early days of the formation of the group. Women reported hearing stories of other SHG members being stoned for membership or SHG women were seen as trouble-makers accused of trying to take over the local council.	"Upper castes say, 'These women attend meetings and visit the panchayat to get money. They are trying to usurp the position of the gowda and take control of the village.'" Mathrani, 2006, South India	Mathrani 2006, Pattenden 2011, Ramachandar 2009	Moderate	Thick description from 2 studies and thin description from 2 study all from South Asia; quality was high for 3 and medium for 1.
	The men used to make comments such as, these women are doing "tamasha" (showing off) and they are going to close down our sangha after a few days. But we did not worry about those comments." Ramachandar, 2009, South India			

Psychological empowerment

In contrast to the quantitative literature, much of the qualitative literature on individual-level empowerment focuses on self-confidence and self-esteem, and suggests women participating in SHGs feel psychologically empowered. The 11 contributing qualitative studies included in this review suggest specific aspects of individual-level change which were experienced by women self-help group members.

Agentic voice: One of the dominant themes from six studies is that women self-help group members reported feeling more capable of speaking in front of others. First, women experienced this by speaking in front of their peers at their group meetings. As groups matured and began to get involved in community development projects, women also talked about feeling capable of speaking in front of others, such as extended families, authorities, and community leaders (Dahal, 2014; Kabeer, 2011; Kilby, 2011; Kumari, 2011; Mathrani & Parodi, 2006; Ramachandar & Peltó, 2009).

Participation in household negotiations: Another emergent theme involved intra-household dynamics, which was mentioned in six studies (Dahal, 2014; Kabeer, 2011; Kumari, 2011; Mercer, 2002; Ramachandar & Peltó, 2009). At first, women reported the process of gaining acceptance from husbands and in-laws to participate in SHGs. Furthermore, women described gaining respect over time from husbands and extended family and becoming decision-makers within their households following their membership in SHGs.

Domestic disputes: Women in eight studies reported how their participation in SHGs had contributed to domestic disputes and violence including both verbal and physical abuse (Dahal, 2014; Kabeer, 2011; Kilby, 2011; Kumari, 2011; Mathrani & Parodi, 2006; Ramachandar & Peltó, 2009; Sahu & Singh, 2012). Women from three studies reported an initial increase in disputes or violence but said that they eventually gained respect from husbands and in-laws by bringing in income to the household. These women also reported fighting less with their husbands (Kumari, 2011; Mathrani & Parodi, 2006; Ramachandar & Peltó, 2009). In two other studies women reported that they experienced a decrease in disputes and conflict between husbands and wives (Knowles, 2014; Sahu & Singh, 2012). In all eight studies, women described how SHG members put social pressure on men to stop beating wives and would show up in groups to support women who had been beaten. The interviewed women felt these activities decreased domestic violence in their communities.

Social empowerment

The literature around empowerment talks about social capital accumulation as a result of participation in SHGs. We found three main themes that emerged within the context of social capital that explain this phenomenon in more detail.

Networking: An important theme discussed by women in seven studies was that not only were women SHG members more confident speaking in front of others, but the women also felt comfortable working with local authorities, village leaders, and law enforcers to make positive changes in their communities (Kabeer, 2011; Kilby, 2011; Knowles, 2014; Kumari, 2011; Mathrani & Parodi, 2006; Pattenden, 2011; Sahu & Singh, 2012). Women's perceptions suggest that these experiences emboldened them to address authorities when a social issue came up or when they needed support for a community development project.

For these women SHG members, networking experiences represented a profound change from being confined to the domestic sphere and speaking only to family and close neighbors. In one group in India, women had to negotiate with formal banking institutions. The women reported that these institutions at first refused to give them loans, but the women went up the chain of authority to the national bank for rural development and their loans were released (Kumari, 2011).

Women suggested that this type of networking was useful in getting small projects completed, and, in four studies, women report that they capitalized on relationships and progressed from holding leadership positions with their groups to holding leadership positions within the community (Knowles, 2014; Kilby, 2011; Kumari, 2011; Mathrani & Parodi, 2006).

Solidarity: Another important theme was the empowerment that came from group solidarity. Women's experiences suggested that knowing that their group is supporting them enabled women to make meaningful decisions and to enact positive change in their lives. This boldness to make change as a result of solidarity was reported with respect to situations within the household or the extended family. Four studies reported on women's perspectives about group solidarity (Dahal, 2014; Kabeer, 2011; Kumari, 2011; Mathrani & Parodi, 2006).

The boldness of women was particularly strong when women talked about how their husbands treated them. Women in three studies (Kabeer, 2011; Kumari, 2011; Mathrani & Parodi, 2006) reported feeling that they now had recourse from the group for husbands who committed such acts as domestic violence and heavy drinking.

Community respect: Similar to this sense of solidarity that was apparent, women reported feeling that being a part of their SHG gave them clout within their communities in five studies (Dahal, 2014; Kabeer, 2011; Kumari, 2011; Sahu & Singh, 2012; Ramachandar & Peltó, 2009). Women described walking confidently through their villages and having the courage to approach authorities in a group whereas before they had not felt this way. The women felt more able to participate in community decision-making, and they felt respected by their peers and their leaders. The women were no longer solely housewives but community actors.

Economic Empowerment

Financial skills and independent decision-making: A theme across seven studies was that women reported feeling empowered by the newness of handling money. Many of the women had never participated in the buying and selling of goods and had never been allowed to manage the household accounting before their SHG membership. With the new access to credit following SHG membership, women were suddenly in the role of the money manager. Although the learning curve was steep for some, most women reported they gained a sense of self-reliance as a result of having access to money, making decisions about buying and selling, and completing transactions with that money (Dahal, 2014; Kabeer, 2011; Kumari, 2011; Maclean, 2012; Mercer, 2002; Ramachandar & Peltó, 2009; Sahu & Singh, 2012.). One interesting finding from two studies was that women stated that they were putting money aside specifically for their daughters' education (Mathrani & Parodi, 2006; Sahu & Singh, 2012).

Financial experience and handling money: Because handling money was a new experience for most women, women in six studies reported feeling unsure about their financial decisions

(Knowles, 2014; Kumari, 2011; Maclean, 2012; Mathrani & Parodi, 2006; Pattenden, 2011; Ramachandar & Peltó, 2009). Some of the SHGs offered training around such topics as income generation and savings. But because women were making decisions in front of their community members, they felt there was a great deal at stake to make sound choices.

In three self-help groups, women reported not feeling prepared to make certain financial decisions related to their individual or group projects (Maclean, 2012; Mathrani & Parodi, 2006; Ramachandar & Peltó, 2009). In one SHG in Bolivia, the women reported that men saw their participation in the SHG as foolish because they were not knowledgeable enough with money to be able to benefit from microfinance services (Maclean, 2012). In another SHG in India, the community was initially discouraging and ready to scorn at any misstep (Mathrani & Parodi, 2006). But in this case, women reported using the public embarrassment to generate greater determination to fix the construction and build a stronger structure. The women reported spending considerable time researching building materials and using their networking skills to find proper builders and building materials in order to redo their community center.

Political empowerment

Catalyzing broader social action: In seven studies, women described their participation in a SHG as a “stepping stone” (Mathrani & Parodi, 2006) toward wider social participation but not necessarily a political act in itself. Participation in SHGs did expose the women to women’s rights through participation in social activities and it did give them political capital through networking (Kumari, 2011; Dahal, 2014) and encouraged them to speak out on political issues such as transparency and accountability (Knowles, 2014; Sahu & Singh, 2012). In addition, women who go on to participate in local village government indicate that participating in SHGs provided the support and grounding for them to be able to take leadership positions in government (Kilby, 2011).

Understanding political context: In three settings, women talked about understanding what they could and could not change in their communities. Women were able to identify barriers to affecting change in their community through even small political acts (Mathrani & Parodi, 2006; Pattenden, 2011; Ramachandar & Peltó, 2009). The context within which groups operated “restricted the capacity for political action” (Pattenden, 2011, p.483). In two other settings, women reported that the gradual acceptance by husbands and community member gave way to broader acceptance and respect, which lent strength to their political efforts (Mathrani & Parodi, 2006; Ramachandar & Peltó, 2009).

But in one case, women reported that changing the status of women in their society was not their priority and not on their stated agenda (Mathrani & Parodi, 2006). In this case it appeared as if women SHG members remained focused on poverty reduction through income-generation and community development—not directly challenging gender norms or women’s status in society. The author of the study reported that things like networking and household decision-making constituted micro-political processes. The author suggests that in this specific case SHG participation did not change the station in life of women: the women were still “tethered to domesticity” and men still regarded women as below them (Kumari, 2011).

In one case, women talked about feeling more aware of their rights but awareness was only an important first step and the women still had a long way to go before women gain property and reproductive rights and the domestic role of women was still primary (Kabeer, 2011).

And as Kabeer (2011) stated when discussing this theme observed in the data from her study:

“In social terms, marriage is still the only conceivable pathway to full adulthood for women, particularly in rural areas. In economic terms, it marks the necessary transition from their dependence on fathers to dependence on husbands and ultimately on sons. On both counts, women had a strong stake in shoring up rather than undermining the institution, however abusive the relationships involved” (p. 519).

Adverse Outcomes

Barriers to Participation: Three studies reported that women talked about barriers to participation including economic and social standing (Dahal, 2014; Mercer, 2002; Mathrani & Parodi, 2006). Specifically, lower class women were excluded from “high class” SHGs and lower caste members were not allowed to mix into upper caste groups due to discrimination. In Tanzania, women reported that wealthier women were more likely than poor women to join SHGs. Women's perceptions suggest that to poor women, the SHG was a status symbol and served to reinforce the idea that wealthier or less poor women had more access to financial services, social capital, and community respect than poorer women (Mercer, 2002). In India, issues of caste and religion came up in terms of participation and groups of the same caste joined together to avoid conflict. But due to limited funding, some groups of the same caste had to wait or did not get funding for their SHG (Mathrani & Parodi, 2006).

Disappointment: Five studies reported that some women felt a degree of disappointment when their groups did not deliver on perceived promises such as solving social problems in their villages such as alcoholism (Mercer, 2002) and challenging cultural norms (Dahal, 2014; Kabeer, 2011). Another source of disappointment occurred when women gained new awareness about rights but were not able to enact them (Kumari, 2011) or when their group took on new responsibilities but in the end did not have the authority or financial power to make changes (Maclean, 2012).

Mistrust and Corruption: In three studies, women reflected on negative experiences about mistrust and corruption of their group or stories about corruption in other groups particularly stories of leaders stealing group funds (Knowles, 2014; Maclean, 2012; Dahal, 2014).

Stigma: Membership in two SHGs had negative associations and women reported facing public shame or discrimination especially during the formation of the groups. This experience of discrimination was reported much less than experiences of increased respect by community member. But importantly, women reported hearing stories of women being stoned for membership (Pattenden, 2011) or SHG women were seen as trouble-makers accused of trying to take over the local council (Mathrani & Parodi, 2012).

4.6. Integrated synthesis

The quantitative synthesis suggests that SHGs have positive effects on women's economic, social, and political empowerment ranging from 0.06–0.41 standard deviations. We did not find quantitative evidence for positive effects of SHGs on psychological empowerment. However, we found that women perceive positive contributions of SHGs to psychological empowerment in the synthesis of the qualitative research. Thus, either the quantitative studies do not adequately measure psychological empowerment or the women's perceptions are biased due to various cognitive biases, such as the fundamental error of attribution or the tendency for people to attribute changes to programs rather than contextual characteristics (White & Phillips, 2012).

The quantitative evidence does not suggest strong adverse impacts of self-help groups on indicators such as disappointment, stigma, or domestic violence, although we were only able to meta-analyze the impact of self-help groups on domestic violence. Findings from the qualitative research suggest that women perceive SHGs as having the potential to reduce domestic violence as a result of some combination of the following: 1) improved economic stability, 2) increased respect of wives by husbands, 3) increased self-confidence of women, 4) exposure to human rights and gender training and 5) enforcement from SHG members to reduce violence within households. These perceptions on domestic violence were one of the strongest themes drawn from eight contributing qualitative studies, although these studies were all conducted in South Asia. However, the quantitative meta-analysis examined the effects of women's self-help groups on attitudes toward domestic violence and the results neither showed evidence for adverse effects of women's self-help groups on attitudes toward domestic violence nor evidence for the potential of SHGs to reduce domestic violence. Thus, we need to be careful in interpreting this result. Nonetheless, our findings certainly do not suggest that there is evidence for increasing the likelihood of domestic violence for SHG participants.

Furthermore, self-help groups may have stronger effects on economic empowerment and women's family-size decision-making power when the self-help group includes a training component. However, we should be careful in the interpretation of this finding because both quantitative and qualitative studies present insufficient details about the contents of the training in SHGs. In the quantitative studies, health education training and training on business and entrepreneurial skills were the most prevalent, but we were not able to distinguish between the effects of different types of training in a meta-analysis because of the limited number of studies.

Although the quantitative analysis did not allow for a rigorous identification of contextual moderators of heterogeneous effects, the qualitative synthesis suggests various reasons for why women do not experience empowerment as a result of women's self-help groups under all circumstances. The first barrier toward an empowering experience resulting from self-help groups that was identified by women SHG members is a barrier to participation in self-help groups. Women SHG members suggest that the poorest of the poor, lower caste members or other marginalized groups may not always have the possibility to participate in SHGs. This perception of women SHG members suggests that the theory of change underlying self-help group interventions we proposed should start even before female participation in economic or livelihood self-help groups. Several assumptions need to be fulfilled before women even start participating in self-help groups. However, we should be

careful in interpreting the result about participation because of the limited potential of qualitative studies to determine causal effects. This finding, nonetheless, reinforces the call of De Hoop and Menon (2014) to more systematically analyze participation in development programs. They argue that:

...while implementing a women's self-help group programme, it would be important to consider the possibility that information about the existence of the programme may not reach potential participants. It is also likely that the women may consider attendance in meetings to have a big opportunity cost. They may have to give up several hours of work on their farm to attend a self-help group meeting. Assumptions about participation may also run counter to what a woman is able to do in her community. Women may have to break gender related social norms to attend this meeting unaccompanied by their spouse or male relative. (De Hoop & Menon, 2014)

The latter argument relates to a different conclusion from the qualitative research. Here, it is interesting to see that women's perspectives suggest that women in Bolivia and Tanzania encounter more resistance from the community when they participate in self-help groups than women in South Asia. Women's perspectives from South Asia suggest that the initial resistance of other community residents to participation of women in self-help groups and the resulting empowerment process fades out after women are self-help group members for a longer amount of time. This finding suggests that the maturity of self-help groups might be an important additional moderator for achieving effects on women's empowerment. With respect to social empowerment, the quantitative evidence suggests this may be true. We found stronger effects of women's self-help groups on women's family-size decision-making power in the context of India, where self-help groups are well-established, than in the context of Ethiopia, where self-help groups are less well-established. However, we have to exercise caution in interpreting this finding because there may be factors that confound this result such as reporting bias and cross-cultural misinterpretation. In addition, the number of studies that discuss backlash from the community is relatively small.

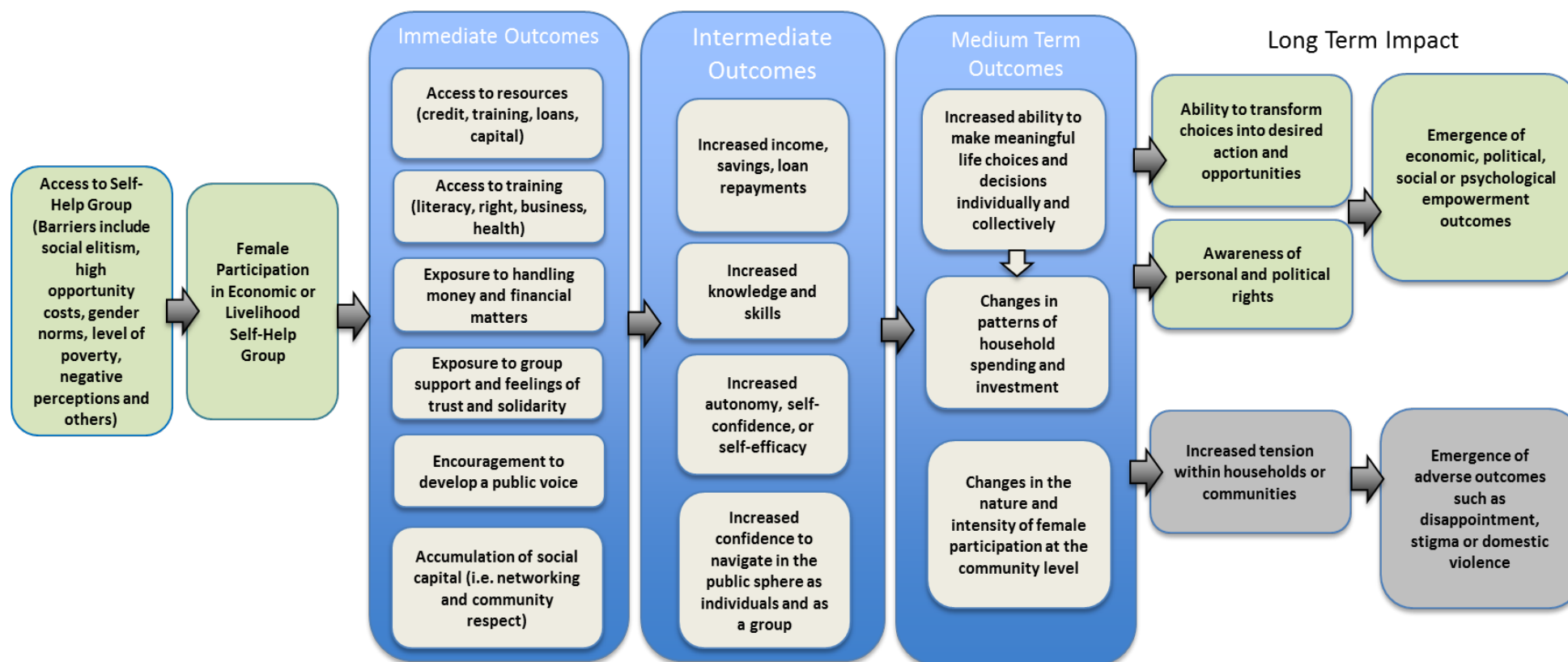
In general, qualitative studies do not give sufficient attention to the identification of causal effects and quantitative studies do not emphasize enough the importance of potential moderators in the design of SHG programs. Too often qualitative studies present information about the empowering experience of women in self-help groups without focusing attention on issues like self-selection in self-help groups. At the same time, our meta-analysis suggests that self-selection in self-help groups complicates counterfactual analysis tremendously. Studies with a high risk of selection bias overestimate the impact of self-help groups relative to studies with a medium or low risk of bias. Furthermore, quantitative studies do not present sufficient detail regarding the specifics of the designs of SHGs. This lack of detail complicates the analysis of moderating effects in the meta-analysis tremendously. The lack of attention for causal identification in the qualitative studies and the lack of detail about the program in the quantitative studies complicate the integration of quantitative and qualitative studies.

Based on the findings discussed above and the relatively small number of quantitative studies that adequately account for selection bias, we argue that, although we are able to determine the average pooled effect of self-help groups on empowerment across studies with reasonable precision, there are not yet enough rigorous quantitative studies about self-help groups that present sufficient details about the program design to answer second-

generation questions with respect to their effectiveness, such as whether self-help groups with a specific training component are more effective than self-help groups that merely provide financial services and group-support. It is still unclear how to organize self-help groups to achieve maximum impacts on women's empowerment. For example, we would need more evidence to understand what types of training result in women's empowerment.

Nevertheless, for other findings, the strength of an integrated mixed-methods review is clearly visible. For example, the quantitative evidence suggests positive effects on various dimensions of empowerment, which the qualitative evidence reinforces with its emphasis on the mechanisms of the underlying the positive effects. Here, the quantitative evidence addresses the attribution question and shows that women's self-help groups have positive causal effects on women's empowerment. The qualitative evidence presents a more nuanced understanding of how these empowerment processes might work. First, women's perspectives suggest that economic empowerment may be stimulated by giving women the opportunity to handle money. Second, women's perspectives indicate that social empowerment may be stimulated by improvements in social networks, community respect, and solidarity among women self-help group members. Third, the integration of the quantitative and qualitative evidence suggests that, although women's self-help groups may stimulate political empowerment, changing the status of women in society is not the main goal of women SHG members. Fourth, women experiences suggest women SHG members are able to speak freely in front of others in contrast to before their membership. These four mechanisms indicate that the original theory of change may miss several intermediate steps in the causal chain. Figure 4.18 depicts a revised theory of change based on our findings.

Figure 4. 18: Revised theory of change



5. Discussion

5.1. Summary of main results

Our results suggest that self-help groups can have positive effects on various dimensions of women's empowerment. We found positive effects, ranging from 0.06–0.41 standard deviations, on economic, and political empowerment, as well as women's family-size decision making power and mobility, which can both be included under social empowerment. However, we did not find evidence for positive effects on psychological empowerment. These findings are based on the results of RCTs and higher quality quasi-experimental studies.

The qualitative synthesis we presented also indicates that women's perspectives suggest that self-help groups contribute positively to their empowerment. The qualitative results showed a more nuanced understanding of how women experience the phenomenon of empowerment after they enter self-help groups. Women's experiences suggested that the positive effects of self-help groups on economic, social, and political, empowerment may run through the channels of familiarity with handling money and independence in financial decision making, solidarity, improved social networks, and respect from the household and other community members. In contrast to the quantitative evidence, the qualitative synthesis of women's perceptions indicate that SHGs may contribute to psychological empowerment.

Our synthesis of women's experiences in SHGs also suggests that while participation in self-help group can initially create tension within households, especially between husbands and wives, in the long term participation in SHGs does not contribute to domestic violence. This finding is in alignment with the lack of evidence for a statistically significant effect of SHGs on the likelihood of domestic violence in our meta-analysis.

The findings on community push-back were mixed and may be context-specific. For example, De Hoop et al. (2014) demonstrated that push-back from conservative community members in India resulted in negative consequences for happiness or subjective well-being for women SHG members, but only in communities with relatively conservative gender norms. Women's perspectives from the qualitative synthesis also present evidence for occasional backlash from other community members. However, our synthesis also suggests that backlash is more prevalent in contexts where SHGs are less well established. Although we have to be cautious in interpreting this result because of the difficulty of establishing causal effects with qualitative research, some of the quantitative studies also presented suggestive evidence that spillovers from self-help groups may benefit the social empowerment of women residents in the community who were not themselves members of self-help groups. These spillovers may be more likely in settings where SHGs are more established.

5.2. Overall completeness and applicability of evidence

A secondary goal of this research was to develop a new theory of change underlying self-help groups using a triangulation of research findings from the quantitative meta-analysis and the qualitative narrative synthesis. As discussed in section 4, the positive effects of self-help groups on various dimensions of women's empowerment indicate that the theory of change we presented at the beginning of this review is at least valid to a certain degree but missed several important steps. A triangulation of the quantitative and qualitative research

findings further indicates that a higher level of group-based support in the form of training might contribute more to women's economic empowerment than the microfinance services of self-help groups. However, we have to be careful in interpreting this result because of the lack of details quantitative studies present about the contents of training in SHGs.

More fundamentally, our research findings also indicate that the original theory of change we presented was not complete. First, the theory of change only started at the stage where women already participate in self-help groups. But, as White (2014) argued, many development programs fail because of the low level of participation in the program or, in other words, the take-up of the program is too low. Our qualitative synthesis suggested that women perceive low participation of the poorest of the poor in self-help group programs. So self-help group programs might currently bring more benefits to a group whose members are not the poorest of the poor. Therefore, we propose to start the theory of change with potential encouragements that might be necessary to stimulate the poorest of the poor to participate in self-help groups. These incentives might be either financial, for example, by offering the opportunity to participate with no savings requirements, or nonfinancial, for example, by stimulating the husbands or mothers-in-law of the poorest of the poor to let their spouses and daughters-in-law participate in self-help group programs.

In addition, our qualitative synthesis suggests that various intermediate outcomes were missing from the original theory of change. First, women's perspectives indicate that SHGs may only contribute to psychological empowerment if women are able to gain a public voice. Second, women's perspectives indicate that women may first need to gain the skill to handle money before women can achieve economic empowerment. Third, women's perspectives suggest that SHGs contribute to social empowerment after women gain respect from community members, which potentially increases the quality of their social networks and improves solidarity among group members. Fourth, women's perspectives about their participation in SHGs suggested that they need to go through various stages of political empowerment, of which only some can be achieved with SHG membership. Women SHG members' perceptions from the qualitative research suggest that women self-help group members only achieved the first stage. In this stage, women became knowledgeable about their rights, but they did not directly challenge women's status in society. Importantly, however, none of the quantitative studies was able to directly test these mechanisms. Thus, we need to remain careful in the interpretation of these results from the qualitative analysis because of the potential of various biases.

With respect to adverse outcomes, our integration of the quantitative and qualitative synthesis suggests that participation in women's self-help groups is not likely to have strong adverse effects on domestic violence. We did not find evidence for positive effects of SHGs on the likelihood of domestic violence. Furthermore, women's perspectives indicate that even if SHGs contribute to domestic violence, this adverse consequence is likely to disappear in the long term.

Finally, the strong heterogeneity in the impact estimates on social empowerment and the wide range of potential mechanisms from the qualitative research indicate the theory of change needs to represent the social and political context within which women are making decisions. For example, women might not choose to become autonomous because this might result in community disapproval. Or women might not choose to participate in a self-help group because then they would no longer have the time required to conduct agricultural

labor. We argue that the considerations discussed previously need to be reflected in the theory of change by adding assumptions along the causal chain from inputs to intermediate to final outcomes. First, women may need to support in introducing the purpose of SHG participation to other household members before they start participating in self-help groups. Second, women need to show demand for the financial and nonfinancial services the self-help group provides, and have sufficient time to participate in the activities of the self-help group.

5.3. Quality of the evidence

The findings of every systematic review depend on the quality of the primary studies on which the review relies. In our case, we believe both the quantitative and the qualitative studies on women's self-help groups suffered from substantial limitations with respect to their quality. However, we also believe that our risk of bias assessment for the quantitative research allowed us to distinguish clearly between the findings of studies with high, medium, and low risk of bias. The meta-analysis indeed showed that studies with a high risk of selection-bias were likely to present biased estimates on the impact of women's self-help groups on women's empowerment. For this reason, we were only able to present a meta-analysis for a small number of studies. We were not able to show strong evidence for heterogeneous effects in a large sample of studies; even though; our analysis presented clear indications for strong heterogeneities in the effect sizes.

In addition, we were not able to present a convincing meta-analysis for the effects of self-help groups on women's psychological and political empowerment. Nonetheless, the results of the meta-analysis for studies with high and medium risks of selection bias presented important evidence of the effects of self-help groups on women's empowerment.

The qualitative evidence also presented important findings with respect to the possible mediators of the effectiveness of women's self-help groups. But the lack of qualitative studies that report the empowering experiences of women in self-help groups directly from women's narratives limited our ability to more fully understand such mediators. Furthermore, several of the qualitative studies suffered from a medium risk of bias.

5.4. Limitations and potential biases in the review process

The limitations of this review are specific to the two types of analyses and appeared in the synthesis process to triangulate the quantitative and qualitative results. In particular, we were not able to triangulate all the research findings with respect to the qualitative synthesis in the quantitative meta-analysis. This was partly because of the small number of studies in the quantitative meta-analysis that could be considered rigorous. More importantly, however, the majority of the potential moderators in the qualitative research were not reported in the quantitative research or insufficient details were provided. Hence, we were not able to estimate the moderating effect of potential moderators identified in the qualitative research. Furthermore, although we were able to assess the moderating effect of training in the quantitative analysis, suggesting that training has positive effects on empowerment, both the quantitative and the qualitative studies did not present sufficient details about the contents of training in SHGs. Thus, we need to remain very careful in the interpretation of this result.

5.4.1 Limitations of quantitative data analysis

Publication bias: The results of our meta-analysis may be vulnerable to publication bias. We tested for the presence of potential publication bias by reporting funnel plots for the effects on women's social (women's family-size decision-making power and mobility) and economic empowerment and reporting the results of the Egger test. From these funnel plots, we concluded that there might be scope for publication bias with respect to the impact estimates of women's self-help groups on women's economic and social empowerment. However, the Egger test did not show formal statistical evidence for publication bias in the impact estimates of women's self-help groups on women's economic or social empowerment. So based on the funnel plots, we can merely say there was potential for publication bias in the studies that focused on women's economic or social empowerment.

In addition, the results of our both our quantitative and qualitative synthesis are heavily based on studies from India and Bangladesh. The external validity of the review may thus be limited to the context of South Asia. At the same time our results may also be most relevant for the context of South Asia because self-help groups are a more popular intervention to stimulate women's empowerment in this region than in other regions of the world.

Missing information: Unfortunately, in this review, we were not able to distinguish among the effects of different self-help group models because studies often did not report sufficient information about the specific model on which they focused. And a wide variety of self-help group models exist across regions. The Indian model was quite different from the model in Bangladesh, and even within India, a wide range of different self-help group models exist. The differences among self-help group models in South Asia and the rest of the world are even larger.

The results of our quantitative synthesis might also be biased due to the exclusion of studies from the meta-analysis for which we were not able to estimate effect sizes. We believe this risk was minimal, however. In general, the findings of the studies we were not able to include in our meta-analysis were consistent with the findings of studies that we were able to include in our meta-analysis. Further, the study findings that were not in line with the findings of the meta-analysis were generally based on studies with a high risk of selection bias. *Data analysis:* Unfortunately, the number of studies with only a low or medium risk of selection bias was limited. Therefore, we were able to convincingly demonstrate the effectiveness of women's self-help groups with respect to social and economic empowerment only. For the effects on psychological and political empowerment we had to mostly rely on a narrative synthesis. In addition, we were not able to convincingly demonstrate the effects of women's self-help groups on women's economic and social empowerment for subgroups using a meta-analysis. For this purpose, we again had to rely on a narrative synthesis. Finally, many studies used different outcome measures to measure the same empowerment domain. Thus, the outcome variables in our meta-analysis may not always measure the same construct. We, however, took this concern seriously as shown by our decision to separately analyze impact estimates on women's family-size decision-making power and women's mobility after our evidence suggested that these two empowerment components cannot be considered part of the same construct. Notwithstanding these limitations, we believe our systematic review presents important evidence with respect to the pooled impact estimate and the heterogeneities in that pooled impact estimate of women's self-help groups on women's economic, social, and less convincingly political and psychological empowerment.

5.4.2 Limitations of qualitative data analysis

Searches: Given the large scope of this review, it is possible that we have missed some articles that may have been relevant. We made a concerted attempt to find all relevant qualitative studies. But we noticed that fewer qualitative evaluations exist and even fewer make it into peer-reviewed publications. Therefore, our search strategy also emphasized the gray literature, including dissertations and unpublished reports. One of the most comprehensive qualitative studies that we included was a dissertation. The value of this piece was further emphasized by the lack of any page limitations, and therefore, the author could include full quotations from female SHG participants and cover many different themes. What appeared in the peer-reviewed literature was less comprehensive, with fewer quotations and less-developed theoretical frameworks. It was unclear if this finding was representative of the lack of strong qualitative studies altogether or if there was a bias in what ends up being published versus the type of research actually conducted. Finally, because of the interdisciplinary nature of the review questions spanning public health, psychology, economics, law, and human rights, it is possible that relevant psychology reports or legal documentation did not meet the inclusion criteria of this review.

Underreporting of adverse outcomes: The included qualitative studies intended to examine changes in empowerment outcomes as stated in their research questions. As a result, qualitative researchers spoke with group members who were willing to talk about their experiences and not with women who did not want to be interviewed, who dropped-out or who did not join SHGs. In addition, researchers did not talk to men or other community members who may have different perspectives about the SHGs. As a result, it may be possible that adverse effects of SHGs were underreported in the qualitative research.

Missing information: Although the authors conducted a thorough quality assessment of each study, there are concerns that descriptions of important methodological processes were missing from many of the qualitative studies. For example, although the data analysis of a study might have appeared rigorous as judged by the results presented, the description of the process of analysis was weak in most studies. In addition, the discussion about the researcher's relationship with the study participants and ethical considerations were either unreported or not examined. These are important parts of any qualitative research and should also be reported in any dissemination of the findings. The risk of bias summary table (4.2) offers a way for readers to assess completeness for themselves.

Data analysis: The meta-ethnographic process attempts to use the included studies much like one would use transcripts in a qualitative analysis. The quality and completeness of the transcripts affects the analysis process and the results. The direct quotations from women in the studies were as close to the raw data as we could get—similar to having access to a dataset in quantitative research. Unfortunately, some studies provided more direct quotations from women than other studies, and the analysis was therefore biased toward studies that included more quotations.

5.4.3 Limitations of the synthesis process

The theory of an integrated mixed-method review is that the two parts of the analysis can inform each other during the analysis process and not just in the conclusions. Therefore, the researchers working on the two parts of the study spent time during the data extraction and analysis phase discussing findings but there were limitations in how much the exchange of

information could impact each analysis. For example, very few concepts that emerged from the qualitative studies could be used in the subgroup analysis of the quantitative studies because of missing data.

We believe that integrated mixed-method reviews that include both quantitative and qualitative research have potential. However, to optimize the learning from integrated mixed-methods reviews, it is important that quantitative researchers integrate the findings of qualitative researchers in their research design and vice versa. Hence, maximizing the potential of integrated mixed-methods reviews would require a more interdisciplinary attitude from both quantitative and qualitative researchers.

5.5. Agreements and disagreements with other studies or reviews

The systematic review found positive significant impacts of self-help groups on empowerment, whereas the systematic reviews that focused on microcredit and microsavings (e.g. Stewart et al., 2012; Vaessen et al., 2014) only found limited evidence for positive effects on economic outcomes. In addition, our quantitative synthesis suggested that self-help group interventions that include a training component may have stronger effects on women's empowerment, particularly economic empowerment and women's family-size decision making power, than self-help groups that do not contain a training component. So although our results presented more positive findings than other systematic reviews with an emphasis on microfinance, we do not believe the results from the different reviews are necessarily contrasting. However, we need to remain very careful in the interpretation of this result because the quantitative studies neither presented sufficient about the training components of SHGs nor about other details of the trainings.

6. Authors' conclusions

6.1. Implications for practice and policy

Our review highlights several important implications for practice and policy related to the rollout and potential impact of SHGs. First, our quantitative evidence suggested positive effects on women's empowerment indicating that self-help groups have the potential to strengthen development outcomes. These findings have important implications for program designers and managers. Thorough program planning and implementation is essential to ensure an optimal number of participants meet frequently. In addition, staff and institutions may consider structures that will ensure the same staff and institutions are accountable to their clients.

The greatest quantitative impacts were found among SHGs where health education, life skills training, and/or other types of information were shared and supported. The additional benefits accrued via group training, such as group sharing, learning, and support. Furthermore, it is important for programs to consider that SHGs offer an important venue to deliver additional services and training. SHGs that are facilitated externally are also more likely to have the resources to provide additional components, such as training. The finding on training might also reflect the success of programs in which more holistic programming is provided as indicated by the qualitative research. However, unfortunately, the quantitative studies do not present details about the contents of the training. Thus, we have to remain careful in interpreting this finding.

One area that has particularly important implications for programs and policy is the qualitative finding that women SHG members perceive low participation of the poorest of the poor in self-help groups. In part, this might be because the poorest of the poor are too financially and/or socially constrained to join self-help groups or to benefit from the financial services most often provided through self-help groups. But other barriers such as class or caste discrimination might also be occurring. Poorer or marginalized women may not feel accepted by groups that are made up of wealthier or better connected community members. It is important for program and policy makers, as well as researchers, to identify ways to build in support and reduce barriers for individual women who want to participate in such groups but who do not have the financial resources or freedoms to join. One enhancement that we have made based on the findings from this review is to start the theory of change related to SHGs with encouragements to stimulate the poorest of the poor to participate in self-help groups. These incentives could be financial, for example, by giving the poorest of the poor the opportunity to participate without a savings requirements, or nonfinancial, for example, by stimulating the husbands or mothers-in-law of the poorest of the poor to let their spouses and daughters-in-law participate in self-help group programs or conducting outreach activities to marginalized groups.

It is important to note that although SHGs overall showed positive impacts, both the quantitative and qualitative evidence showed there was much heterogeneity across program designs and the effectiveness of programs. This finding indicates that context matters. The types of specific program components, and the likely impacts, depend on the overarching social, cultural, political, and economic context from a national level down to a very local level. As new programs are implemented in different contexts, and as more nascent groups become more established, it is critical that program designs are tailored to the local settings

in ways that allow them to evolve over time. Such consideration may include conducting community readiness activities, performing more comprehensive outreach to marginalized groups even within small communities, and included some form of advocacy training so that women might address change beyond the individual level and towards overcoming structural barriers to empowerment. This review has shown that one-size does not fit all, and while there is a need to take best practices across programs for implementation, this needs to be done in a flexible way to adapt programs most successfully for the greatest impact in women's lives.

6.2. Implications for research

This review has several implications for research. First, the synthesis of the quantitative evidence suggests there is a need for more rigorous quantitative studies that can correct for selection bias, spillovers, and the difficulties of measuring empowerment. The quantitative synthesis indicated that studies that did not adequately account for selection bias overestimated the impact of self-help groups on empowerment. Furthermore, the qualitative synthesis suggested that the current measurements of empowerment in the quantitative studies might not reliably capture all dimensions of empowerment. Whereas the quantitative measures are useful in understanding certain aspects of the impact of self-help groups on empowerment, the qualitative studies show us more nuanced ideas about how to measure the lived experience of empowerment. In both cases (quantitative and qualitative studies), researchers need to describe more fully the various components of the interventions/programs being studied, so outcomes and findings can be understood and interpreted against the specifics of the program components. Greater detail in the description of the program design will help in determining moderating factors in the design of SHGs. In addition, future research could draw on mixed-method strategies to develop and test new rigorous measures of empowerment.

Second, there is a need for more research focused on examining the impact of economic self-help groups on women's empowerment using mediator and moderator analysis to further understand the pathways or mechanisms through which SHGs impact empowerment. In addition, there may be other pathways not examined in this review that lead to empowerment that can be rigorously measured (or measure development embarked on) for inclusion in future studies that examine the impact of women's SHGs on empowerment. Potential mediators/moderators of interest include indicators of mental health, relationship power, community-level respect, social capital, and social solidarity. In addition, other important mediators may include an understanding of whether women who participate in SHGs have male partners who experience shifts in gender-related attitudes in the direction of more gender equality as measured by the gender equitable man scale (Pulerwitz et al., 2008). Future research can examine if and how men's attitudinal shifts impact, positively or negatively, women's empowerment. Furthermore, the effects of these complex interventions take time to influence both mediators and outcomes; thus, longer follow-up periods are needed in future research to understand fully both the long-term impacts of SHGs and the factors that support the maintenance of empowerment.

Because women's self-help programs are implemented across many different regions of the world, it is also critical for researchers to not assume that an intervention that works in one place should be replicated elsewhere. In short, as alluded to, nuanced modifications of

programs and sensitivity to local cultural norms are needed in future program design and in the evaluation of program impacts.

Another interesting dimension of our review, where we were not able to make definitive conclusions, is the effectiveness of SHGs that integrate components other than economic ones (skills-building, reduced family size, reproductive health) and whether these “integrated programs” result in more social, psychological, political, or economic empowerment for women.

7. Acknowledgments

We would like to thank 3ie for funding this study. We would also like to acknowledge our advisory group including Reema Nanavaty of SEWA and Shahid Vaziralli from the Center for Microfinance, search strategist Tara Horvath and research assistants, Tra Truong, Julie Weiland and Keely Molina Johnson. Furthermore, we would like to thank Hugh Waddington and John Evers for very thoughtful comments and suggestions on earlier versions of this review and David Wilson for support in the calculation of the effect sizes.

References

Included Quantitative Studies (review objective 1)

- Ahmed, S. (2005). "Intimate partner violence against women: Experiences from a woman-focused development programme in Matlab, Bangladesh," *J Health Popul Nutr*, Mar. 23(1): 95–101.
- Bali Swain, R. B. & Wallentin, F. Y. (2009). "Does microfinance empower women? Evidence from self-help groups in India." *International Review of Applied Economics*, 23(5): 541–556.
- Banerjee, A., Duflo, E., Glennerster, R. & Kinnan, C. (2015). "The miracle of microfinance? Evidence from a randomized evaluation." *American Economic Journal: Applied Economics*, 7(1): 22-53.
- Coleman, B. E. (2002). "Microfinance in Northeast Thailand: Who benefits and how much?" ERD Working Paper Series no. 9, Economics and Research Department, Asian Development Bank.
- De Hoop, T., van Kempen, L., Linssen, R., & van Eerdewijk, A. (2014). Women's Autonomy and Subjective Well-Being: How Gender Norms Shape the Impact of Self-Help Groups in Odisha, India. *Feminist Economics*, 20(3), 103-135.
- Deininger, K. & Liu, Y. (2013). "Economic and Social Impacts of an Innovative Self-Help Group Model in India." *World Development*, 43: 149–163.
- Desai, R. M. & Joshi, S. (2012). "Collective action and community development: Evidence from women's self-help groups in rural India." Available at http://www.isid.ac.in/~pu/seminar/31_07_2012_Paper.pdf.
- Desai, J. & Tarozzi, A. (2011). "Microcredit, family planning programs, and contraceptive behaviour: Evidence from a field experiment in Ethiopia." *Demography*, 48: 749–782.
- Garikipati, S. (2008). "The impact of lending to women on household vulnerability and women's empowerment: evidence from India." *World Development*, 36(12): 2620–2642.
- Garikipati, S. (2012). "Microcredit and women's empowerment: Through the lens of time-use data from rural India." *Development and Change*, 43(3): 719–750.
- Holvoet, N. (2005). "The impact of microfinance on decision-making agency: Evidence from South India." *Development and Change*, 36(1): 75–102.
- Husain, Z., Mukherjee, D. & Dutta, M. (2010). "Self-help groups and empowerment of women: Self-selection or actual benefits?" MPRA Paper No. 20765, Munich Personal RePEc Archive.
- Kim, J., Ferrari, G., Abramsky, T., Watts, C., Hargreaves, J., Morison, L., Phetla, G., Porter, J. & Pronyk, P. (2009). "Assessing the incremental effects of combining economic and health interventions: IMAGE study in South Africa." *Bulletin of the World Health Organisation*, 87: 824–832.
- Mahmud, S. (1994). "The role of women's employment programmes in influencing fertility regulation in rural Bangladesh." *The Bangladesh Development Studies*, XXII (2, 3): 93–119.

Mukherjee, A. K. & Kundu, A. (2012). "Microcredit and women's agency: A comparative perspective across socioreligious communities in West Bengal, India." *Gender Technology and Development*, 16(1): 71–94.

Nessa, T., Ali, J. & Abdul-Hakim, R. (2012). "The impact of microcredit program on women's empowerment: Evidence from Bangladesh." *OIDA International Journal of Sustainable Development*. 03(09): 11–20.

Osmani, L. N. K. (2007). "A breakthrough in women's bargaining power: The impact of microcredit." *Journal of International Development*, 19: 695–716.

Pitt, M. M., Khandker, S. R. & Cartwright, J. (2006). "Empowering women with micro-finance: Evidence from Bangladesh." Available at http://www.brown.edu/research/projects/pitt/sites/brown.edu.research.projects.pitt/files/uploads/EDCC2006_0.pdf.

Rosenberg, M. S., Seavey, B. K., Jules, R. & Kershaw, T. S. (2011). "The role of a microfinance program on HIV risk behaviour among Haitian women." *AIDS and Behavior*, 15: 911–918.

Sherman, S. G., Srikrishnan, A. K., Rivett, K. A., Liu, S., Solomon, S. & Celentano, D.D. (2010). "Acceptability of a microenterprise intervention among female sex workers in Chennai, India." *AIDS and Behavior*, 14: 649–657.

Steele, F., Amid, S. & Naved, R. T. (1998). "The impact of an integrated micro-credit program on women's empowerment and fertility behaviour in rural Bangladesh." Available at <http://www.popcouncil.org/uploads/pdfs/wp/115.pdf>.

Swendeman, D., Basu, I., Das, S., Jana, S. & Rotheram-Borus, M. J. (2009). "Empowering sex workers in India to reduce vulnerability to HIV and sexually transmitted diseases." *Social Science & Medicine*, 69(8): 1175–1166.

Included Qualitative Studies (review objective 2)

Dahal, S. (2014). "A Study of women's self-help groups (SHGs) and the impact of SHG participation on women empowerment and livelihood in Lamachur village of Nepal." Masters Thesis, Department of International Environment and Development Studies, Norwegian University of Life Sciences, Norway.

Kabeer, N. (2011). "Between Affiliation and Autonomy: Navigating Pathways of Women's Empowerment and Gender Justice in Rural Bangladesh." *Development and Change* 42(2): 499–528.

Kilby, P. (2011). "NGOs in India: The challenges of women's empowerment and accountability." Routledge Contemporary South Asia Series.

Knowles, G.E. (2014). "An Examination of Microfinance Self Help Groups and the Poorest of the Poor Women in Tamil Nadu, India." Doctoral Dissertation, School of Accountancy, Queensland University of Technology, Brisbane, Australia.

Kumari, K. B. V. (2011). Microcredit as a poverty alleviation strategy, women's empowerment and gender relations. Dissertation thesis, Graduate School – New Brunswick Rutgers, The State University of New Jersey, New Brunswick, New Jersey.

Maclean, K. (2012). "Banking on women's labour: Responsibility, risk and control in village banking in Bolivia." *Journal of International Development*, 24: S100–S111.

Mathrani, V. & Parodi, V. (2006). "The Sangha Mane: The translation of an internal need into a physical space." *Indian Journal of Gender Studies*, 13: 317–349.

Mercer, C. (2002). "The disclosure of *Maendeleo* and the politics of women's participation on Mount Kilimanjaro." *Development and Change*, 33: 101–127.

Pattenden, J. (2011). "Social protection and class relations: Evidence from scheduled caste women's associations in rural South India." *Development and Change*, 42(2): 469–498.

Ramachandar, L. & Pelto, P. J. (2009). Self-help groups in Bellary: Microfinance and women's empowerment. *The Journal of Family Welfare*, 55(2): 1–16.

Sahu, L., & Singh, S. K. (2012). A Qualitative Study on Role of Self Help Group in Women Empowerment in Rural Pondicherry, India. *National Journal of Community Medicine*, 3(3):473-9.

Excluded Studies

Quantitative

Ackerly B (1995). "Testing the tools of development; credit programmes, loan involvement and women's empowerment" *IDS Bulletin* 26:3 (56-68).

Ashburn K, Kerrigan D, Sweat M (2008). "Micro-credit, Women's Groups, Control of Own Money: HIV-Related Negotiation Among Partnered Dominican Women" *AIDS Behavior* 12:296-403.

Banerjee N (2004). "Nari Bikash Sangha: Towards Empowerment" *Indian Journal of Gender Studies* 11:2(179-203).

Bushamuka V, de Pee S, Talukder A, Kiess L, Panagides D, Taher A, Bloem M. (2005). "Impact of a homestead gardening program on household food security and empowerment of women in Bangladesh" *Food and Nutrition Bulletin* 26:1.

Chandra P and Sinha A (2010). "Performance and Sustainability of Self-help Groups in India: A gender Perspective" *Asian Development Review* 27:1.

Deininger K and Lui Y (2009). "Longer-Term Economic Impacts of Self-Help Groups in India. World Bank Policy Research Working Paper 4886, Development Research Group.

Euser S, Souverein D, Rama P, Gowda N et al. (2012). "Pragati: An empowerment program for female sex workers in Bangalore, India." *Global Health Action* 5:19279.

Feigenberg, Benjamin, Erica M. Field, and Rohini Pande. (2010). Building Social Capital Through Microfinance. HKS Faculty Research Working Paper Series, RWP10-019, John F. Kennedy School of Government, Harvard University.

Lokhande M (2013). "Micro Finance for Women's empowerment - A Study of Self-help Groups-Bank Linkage Programme." *International Center for Business Research* 2: Apr 2013.

- Madheswaran, S. and Amita Dharmadhikary (2001). "Empowering Rural Women through Self-help Groups: Lessons from Maharashtra Rural Credit Project", *Indian Journal of Agricultural Economics*, 56:3.
- Mansuri, B.B. (2010). Micro financing through Self Help Groups – A Case Study of Bank Linkage Programme of NABARD. *APJRB*, 1 (3), 141 – 150.
- Mayoux L (2005). "Participatory Action Learning Systems (PALS): Impact Assessment of Civil Society Development and Grassroots-based Advocacy in Anandi, India" *Journal of International Development* 17(211-242).
- Murthy, PSR. (2012). "Economic Empowerment of Rural Women by Self Help Group through Micro Credit." (December 22, 2012). Available at SSRN: <http://ssrn.com/abstract=2193045>
- Odutolu, O (2003). "Economic Empowerment and Reproductive Behaviour of Young Women in Osun State, Nigeria" *African Journal of Reproductive Health* 7:3(92-100).
- Oosterhoff P, Anh NT, Yen PN, Wright P, Hardon A (2008). "HIV-positive mothers in Viet Nam: using their status to build support groups and access essential services" *Reproductive Health Matters* 16(32):162–170.
- Panda, D.K, (2009). 'Participation in the Group Based Microfinance and its Impact on Rural Households: A Quasi-experimental Evidence from an Indian State, *Global Journal of Finance and Management*, Vol. 1 (2) pp. 171-183
- Parajuli D, Acharya G, Chaudhury N and Thapa B (2012). "Impact of Social Fund on the Welfare of Rural Households Evidence from the Nepal Poverty Alleviation Fund" *The World Bank Policy Research Working Paper* 6042.
- Premaratne SP, Senanayahe SMP and Warnasuriya M (2012). "Empowerment of Women through Self Help Groups (SHGs): A Study of a SHG Microfinance Project in Sri Lanka." *UEH Journal of Economic Development*, No. 210, February 2012, University of Economics, Vietnam. pp 17-36.
- Pronyk P, Harpham T, Busza J, Phetla G et al. (2008). "Can social capital be intentionally generated? A randomized trial from rural South Africa" *Social Science & Medicine*, 67:1559–1570.
- Pucho, K (2008). "Evaluation on Women Self-Help Group in the district of Mokokchung, Mon & Phek." Kohima: Directorate, State Institute of Rural Development, India.
- Reddy, C.S. & Manak, S. (2005). *Self Help Groups: A Keystone of Microfinance in India – Women's empowerment and Social Security*. Mahila Abhivruddhi Society, Andhra Pradesh.
- Sabhlok S (2011). "Development and Women: The Role of Trust in Self-Help Groups" *Indian Journal of Gender Studies*, 2011 18: 241.
- Saha S, Annear PL, Pathak S (2013). "The effect of Self-Help Groups on access to maternal health services: evidence from rural India" *International Journal for Equity in Health*, 2013, 12:36.
- Sinha A, Parida PC, Baurah P (2012). "The impact of NABARD's Self Help Group-Bank Linkage Programme on Poverty and Empowerment in India" *Contemporary South Asia*, 20:4(487-510).

Sinha S and Pastakia A (2004). "Evaluation Report: GOI-UNDP Community-based Pro-poor Initiative Programmes." United Nations Development Programme.

Suguna B (2006). "Empowerment of Rural Women Through Self Help Groups" Discovery Publishing House, New Delhi, India.

Swain, R. B., & Varghese, A. (2009). Does Self Help Group participation lead to asset creation? *World Development*, 37(10), 1674-1682.

Teshome E (2012). "The Role of Self-Help Voluntary Associations for Women's empowerment and Social Capital: The Experience of Women's Iddirs (Burial Societies) in Ethiopia" *Journal of Community Health* 73:706-714.

UNFPA (2006). "Exploring Linkages: Women's Empowerment, Microfinance and Health Education."

Urquieta J, Tepichin-Valle A, Tellez-Rojas M (2009). "Poverty and Gender Perspective in Productive Projects for Rural Women in Mexico: Impact Evaluation of a Pilot Project" *Evaluation Review* 33:1(27-53).

Uys LR, Bhengu BR and Majumdar B (2006). "The use of PHC clinic-based women's groups for financial empowerment in a rural area" *Curationis*, March 2006.

Van Kempen L (2009). "The 'Downside' of Women's empowerment in India: An Experimental Inquiry into the Role of Expectations" *Social Indicators Research* 94:465–482.

Qualitative

Ahmed, S. (2011). "Development Innovations through Entrepreneurial Microfinance and the Attempt to Achieve the United Nations Millennium Development Goals in Bangladesh." *Advances in Social Work*, 13:2(358-374).

Apantaku S. (2008). "Assessment of Community Empowerment and Women Involvement Components in National Fadama 11 Project in Ogun State, Nigeria." *Journal of Social Development in Africa*, 23:2(107-127).

Barry K (2010). "Women's empowerment and Community Development through Ecotourism" STI Graduate Institute/STI Study Abroad Capstone Collection. Paper 2579.

Bhatt N and Tang Shui-Yan T (2001). "Delivering Microfinance in Developing Countries: Controversies and Policy Perspectives" *Policy Studies Journal*, 29:2(319-333).

Bhengu BR (2010). "An investigation into the level of empowerment of rural women in the Zululand district of KwaZulu-Natal province of South Africa" *Curationis* 33(2):4-12.

Biradavolu MR, Burris S, George A, Jena A and Blankenship K (2009). "Can sex workers regulate police? Learning from an HIV prevention project for sex workers in southern India" *Social Science & Medicine* 68 (2009) 1541–1547.

Cheston, S. Kuhn, L. (2002). "Empowering Women through Microfinance." UNIFEM, New York, NY.

Faraizi, A. Rahman T, McAllister J. (2011). "Microcredit and Women's Empowerment: A Case Study of Bangladesh." *Routledge Contemporary South Asia*, New York, NY.

Ghadoliya MK (2003). "Empowering Women Through Self -Help Groups: Role Of Distance Education." *Commonwealth of Learning*.

Gibb, S. (2008). "Microfinance's Impact on Education, Poverty, and Empowerment: A Case Study from the Bolivian Altiplano." Institute for Advanced Development Studies Development Research Working Paper Series, No. 04/2008.

Guerin, I. (2006). "Women and Money: Lessons from Senegal." *Development and Change* 37:3(549–570).

Hoodfar H (2010). "Health as a Context for Social and Gender Activism: Female Volunteer Health Workers in Iran" *Population and Development Review* 36(3): 487–510.

Ialsm, N., Ahmed, E., Chew, J., D'Netto, B. (2011). "Empowerment Factors Of Rural Women Through Homestead Poultry Rearing In Bangladesh." International Business Research Conference, World Business Institute Australia, Melbourne, Australia.

Jerinabi U. (2008). "Micro enterprises for women: competitiveness, challenges, and prospects for new global environment." *Discovery Publishing House*. New Delhi, India.

Kim, J. et al. (2007). "Understanding the Impact of a Microfinance-Based Intervention on Women's Empowerment and the Reduction of Intimate Partner Violence in South Africa." *American Journal of Public Health*. 97:10.

Kuttab E (2010). "Empowerment as Resistance: Conceptualizing Palestinian women's empowerment" *Development* 53:2(247-253).

Lokhande MA (2008). Socio- Economic Impact of Micro Financing through Self- Help Groups in Marathwada Region. *The Indian Journal of Commerce* vol. 61, No.4, October-December 2008: 151-164.

Lombe M, Newransky C and Kayser K (2012). "Exploring Barriers to Inclusion of Widowed and Abandoned Women through Microcredit Self-Help Groups: The Case of Rural South India" *Journal of Sociology & Social Welfare*, 39:2.

Mayoux L. (2001). "Tackling the Down Side: Social Capital, Women's Empowerment and Micro-Finance in Cameroon." *Development and Change*, 23:421-450.

Mayoux L (2000). "Micro-finance and the empowerment of women: A review of the key issues" *Social Finance Working Paper* No. 23.

Meena, H., & Meena, K. (2011). *Women's empowerment and self-help group: Gender disparity and democratic safety*. Jaipur: Aadi Publications.

Moyle T, Dollard M, Biswas S (2006). "Personal and Economic Empowerment in Rural Indian Women: A Self-Help Group Approach." *International Journal of Rural Management*, 2(2): 245-266.

Ndinda C (2009). "'But now I dream about my house': women's empowerment and housing delivery in urban KwaZulu-Natal, South Africa." *Development Southern Africa* 26:2.

Nguyen TA, Oosterhoff P, Ngoc Y, Wright P and Hardon A (2009). "Self-Help Groups Can Improve Utilization of Postnatal Care by HIV-Infected Mothers." *J Assoc Nurses AIDS Care*. Mar-Apr; 20(2):141-52.

- Nkosi A (2003). "An Evaluation of a Women's empowerment Life Skills Programme in an Informal Settlement" Dissertation submitted to Rand Afrikaans University, August 2003.
- Noreen U, Omran R Zaheer A and Iqbal Saif M (2011). "Impact of Microfinance on Poverty: A Case of Pakistan." *World Applied Sciences Journal* 12:6(877-883).
- Norwood, C. (2004). "Microeconomic Schemes in Rural Ghana: A Mixed-Method Evaluation of Microcredit Membership on Women's Empowerment, and Family Planning Practices." A Dissertation Presented to the Faculty of The Graduate College at the University of Nebraska.
- Patel S (2014). Empowerment Of Women Through Self Help Groups A Comparative Study Of Tribal And Non Tribal Districts In Orissa. Unpublished Dissertation from University of Delhi, India.
- Panwar M and Kumar J (2010). "Self Help Groups (Shgs) Of Women In Haryana: A Social Work Perspective" *OIDA International Journal of Sustainable Development* 05: 10.
- Pronyk, P. et al. (2008a). "A combined microfinance and training intervention can reduce HIV risk behaviour in young female participants." *AIDS* 2008, 22:1659–1665.
- Pronyk, P. et al. (2008b). "Can social capital be intentionally generated? A randomized trial from rural South Africa." *Social Science and Medicine*. 67:1559–1570.
- Rahman S (2011). "Consumption Difference Between Microcredit Borrowers and Non-borrowers: A Bangladesh Experience." *The Journal of Developing Areas*, 43: 2(313-326).
- Reza-Paul, S. et al. (2012). "Sex worker-led structural interventions in India: a case study on addressing violence in HIV prevention through the Ashodaya samithi collective in Mysore." *Indian Journal of Medical Research*. 135: 98-106.
- Roger U, Rutten A, Frahsa A, Abu-Omar K and Morgan A (2011). "Differences in individual empowerment outcomes of socially disadvantaged women: effects of mode of participation and structural changes in a physical activity promotion program." *Int J Public Health* 56:465–473.
- Salway, S., Jesmin, S. and Rahman, S. (2005). "Women's Employment in Urban Bangladesh: A Challenge to Gender Identity?" *Development and Change* 36(2): 317–349
- Sabhlok, SG., (2011). "Development and Women: The Role of Trust in Self-Help Groups." *Indian Journal of Gender Studies*. 18: 241.
- Sarojani B, (2006). "Women's empowerment through Self-Help Groups", Ambala Cantt: Associated Publishers, 2006.
- Sharma G (2014). Microfinance through SHGs and the Empowerment of Women: A case study of some selected SHGs of Sikkim. *Asian Journal of Multidisciplinary Studies*, 2014, 2:7.
- Shylendra, H.S. (1999). "Promoting Women's Self-Help Groups Lessons from an action research project of IRMA" Anand, India. Working paper no.121.
- Somé, B. (2013). "'Hot Money': Gender And The Politics Of Negotiation And Control Over Income In West African Smallholder Households." *Africa* 83:2(251-269).

Sotshongaye, A & Moller, V. (2000). "We want to live a better life like other people': self-assessed development needs of rural women in Ndwedwe, KwaZulu-Natal." *Development Southern Africa*. 17:1(117-134).

Ssewamala, F., Ismayilova, L., McKay M, Sperber E, Bannon Jr., W and Alicea S. (2010). "Gender and the Effects of an Economic Empowerment Program on Attitudes Toward Sexual Risk-Taking Among AIDS-Orphaned Adolescent Youth in Uganda." *J Adolesc Health*. 46(4): 372–378

Torri M & Martinez, A. (2011). "Gender Empowerment and Equality in Rural India: Are Women's Community-Based Enterprises the Way Forward?" *Journal of International Women's Studies*. 12:1(157-176).

Tupe N and Kamble J (2013). "Micro finance & Micro credit As a Solutions for the Income, Consumption and Investment of Women through SHG in Maharashtra" *Scholarly Research Journal for Interdisciplinary Studies*, 1:4.

Vijayanthi KN (2009). "Women's Empowerment through Self-help Groups: A Participatory Approach" *Indian Journal of Gender Studies* 2002 9: 263.

Additional References

The African Women's Communication and Development Network. (2009). "The African Women's Regional Shadow Report on Beijing Plus 15." Retrieved from http://www.un-ngls.org/IMG/pdf_1272966511_196.200.26.62_Africa_NGO_Report-_Beijing_15_FINAL-ENG.pdf.

Ahmed, S. & Chowdhury, M. (2001). "Micro-credit and emotional well-being: Experience of poor rural women from Matlab, Bangladesh." *World Development* 29(11): 1957–1966.

Amaza, P.S., Kwagbe, P.V. & Amos, A.A. (1999). "Analysis of women participation in agricultural cooperatives: Case study of Borno State, Nigeria." *Annals of Borno* 15/16: 187–196.

Angrist, J. D. & Pischke, J-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Atkins, S., Lewin, S., Smith, H., Engler, M., Fretheim, A. & Volmink, J. (2008). "Conducting a meta-ethnography of qualitative literature: Lessons learnt." *BMC Medical Research Methodology* 8: 21.

Attanasio, O., Augsburg, B., De Haas, R., Fitzsimons, E. & Harmgart, H. (2015). "The impacts of microfinance: Evidence from joint-liability lending in Mongolia." *American Economic Journal: Applied Economics*, 7(1): 90-122

Augsburg, B., De Haas, R., Harmgart, H. & Meghir, C. (2015). "The impacts of microcredit: Evidence from Bosnia and Herzegovina." *American Economic Journal: Applied Economics*, 7(1): 183-203

Barker, G. & Schulte, J. (2010). "Engaging men as allies in women's economic empowerment: strategies and recommendations for CARE country offices." Available

at: http://www.google.com/search?site=&source=hp&q=gary+barker+men+microfinance&og=gary+barker+men+microfinance&gs_l=hp.3..33i21.258.3187.0.3299.28.18.0.0.0.0.294.190.8.1i8j3.12.0....0...1c.1.19.hp.q-X1pgat6is.

Bland, J. M., & Altman, D. G. (2000). "Statistics Notes: The Odds Ratio." *British Medical Journal*, 320: 1468.

Borenstein, M., Hedges, L.V., Higgins, J.P.T., and Rothstein, H. (2009). *Introduction to Meta-Analysis*. Chichester: Wiley.

Brickley, D. B., Almers, L., Kennedy, C. E., Spaulding, A. B., Mirjahangir, J., Kennedy, G.E., Packel, L., Osborne, K., Mbizvo, M., Collins, L. (2011). "Sexual and reproductive health services for people living with HIV: A systematic review." *AIDS Care*. Mar. 23(3): 303–314.

Bohren MA, Vogel JP, Hunter EC, Lutsiv O, Makh SK, Souza JP, et al. (2015). The Mistreatment of Women during Childbirth in Health Facilities Globally: A Mixed-Methods Systematic Review. *PLoS Med* 12(6): e1001847. doi:10.1371/journal.pmed.1001847

Brody, C, Dworkin S, Dunbar M, Murthy P. (2014). "The Effects Of Economic Self-Help Group Programs On Women's Empowerment: A Systematic Review Protocol," The Campbell Library, November 2014.

Campbell Systematic Reviews (2014). "Campbell Collaboration Systematic Reviews: Policies and Guidelines: Version 1.0." Available at http://campbellcollaboration.org/lib/download/3308/C2_Policies_Guidelines_Version_1_0.pdf

Chasnow, M. (2008). "What is microfinance 'plus'?" *India Development Blog: What Is Microfinance 'Plus'?* N.P., Oct. 30. Accessed Mar. 6, 2013.

Coleman, B. E. (1999). The impact of group lending in Northeast Thailand. *Journal of development economics*, 60(1), 105-141.

Colvin, C. (2011). "Executive summary report on the impact of Sonke Gender Justice's one man can campaign in Limpopo, Eastern Cape, and Kwa-Zulu Natal, South Africa." Available at www.genderjustice.org.za.

Crepon, B., Devoto, F., Duflo, E. & Pariente, W. (2015). "Estimating the impact of microcredit on those who take it up: Evidence from a randomized experiment in Morocco." *American Economic Journal: Applied Economics*, 7(1): 123-250.

Critical Appraisal Skills Programme (CASP) (2013). Collaboration for Qualitative Methodologies. Available from <http://www.casp-uk.net/wp-content/uploads/2011/11/CASP-Qualitative-Research-Checklist-31.05.13.pdf>.

Dalal, K. (2011). "Does economic empowerment protect women from intimate partner violence?" *J Inj Violence Res.*, Jan. 3(1): 35–44.

Datta, P. B. & Gailey, R. (2012) "Empowering women through social entrepreneurship: Case study of a women's cooperative in India." *Entrepreneurship Theory and Practice*. doi:10.1111/j.1540-6520.2012.00505.x

De Hoop, T. & Menon, R. (2014) "Making participation count." Available at <http://blog.3ieimpact.org/making-participation-count/>.

De Hoop, T. (2012). *How Context Matters for Development Effectiveness: A Study into Social Norms and Heterogeneous Impacts*. Enschede, The Netherlands: Ipskamp Publishers.

Denison, J. A., O'Reilly, K. R., Schmid, G.P., Kennedy, C.E., Sweat M.D. (2008). "HIV voluntary counseling and testing and behavioral risk reduction in developing countries: A meta-analysis, 1990-2005." *AIDS and Behavior*, 12: 363–373

Dunbar, M. S. & Mudekanye, I. (Under review). *Shaping the Health of Adolescents in Zimbabwe (SHAZ!): Results of a Phase II Randomized Controlled Trial of an Economic Livelihoods Intervention for HIV Prevention among Adolescent Female Orphans*. In *Justice and in Health (Women's Health and Empowerment Center of Excellence textbook*. Berkeley: UC Press.

Dunbar M, Maternowska MC, Kang MS< Laver SM, Medekenye-Mahala I, Padian NS (2010). "Findings from SHAZ!: a feasibility study of a microcredit and life-skills HIV prevention intervention to reduce risk among adolescent female orphans in Zimbabwe." *J Prev Interv Community*. 2010 Apr;38(2):147-61.

Duvendack, M., Palmer-Jones, R., Copestake, J. G., Hooper, L., Loke, Y. & Rao, N. (2011). *What Is the Evidence of the Impact of Microfinance on the Well-Being of Poor People?* London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Dworkin, S. L. (Forthcoming). *Men at Risk: Gender Relations and HIV Prevention*. New York: NYU Press.

Dworkin, S. L., Dunbar, M., Krishnan, S., Hatcher A., Sawires, S. (2011). Uncovering tensions and capitalizing on synergies in violence and HIV programming. *American Journal of Public Health*, 101: 995–1003.

Dworkin, S. L., Dunbar, M. S., Krishnan, S., Hatcher, A. M. & Sawires, S. (2010). "Uncovering tensions and capitalizing on synergies in HIV/AIDS and antiviolenence programs." *Am J Public Health*, 191106v1.

Dworkin, S. L. & Hatcher A. (2012). "Microfinance and health: What works, where, when, why/how? An integrated systematic and critical review of the global evidence base." Prepared for the Centers for Disease and Control and Common Health Action.

Eyben, R., Kabeer, N. & Cornwall, A., (2008). "Conceptualising Empowerment and the Implications for Pro Poor Growth." Paper prepared for the DAC Poverty Network by the Institute of Development Studies, Brighton.

Food and Agriculture Organization (FAO) (2011). "The Role of Women in Agriculture" ESA Working Paper No. 11-02.

Fleming, P., Lee, J. & Dworkin, S. L. (2014). "'Real Men Don't': Constructions of Masculinity and Inadvertent Harm in Gender-Transformative Health Programming." *American Journal of Public Health*, 104: 1029–1035.

Harden, A. (2010). "Mixed-methods systematic reviews: integrating quantitative and qualitative findings." The National Center for the Dissemination of Disability Research (NCDDR), Technical Brief, No. 25.

Higgins, J. P. T. & Green, S., eds. (2011). *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0 (updated March 2011). The Cochrane Collaboration Available at <http://www.cochrane-handbook.org/>.

Hombrados, J. G., and Waddington, H., (2012). *Internal Validity in Social Experiments and Quasi-Experiments: An Assessment Tool for Reviewers*. Mimeo: 3ie.

IFAD. (2003). "Empowering Women Through Self-Help Microcredit Programmes." *Bulletin on Asia-Pacific Perspectives* 2002/03.

International Finance Corporation (IFC) (2014). "Women-Owned SMEs: A Business Opportunity for Financial Institutions." Washington, D.C.

Jakimow, T. & Kilby, P., (2006). "Empowering women: a critique of the blue-print for self-help groups in India." *Indian Journal of Gender Studies* 13(3): 375–400.

Janssens, W. (2011). "Externalities in program evaluation: The impact of a women's empowerment program on immunization." *Journal of the European Economic Association*, 9(6): 1082–1113.

Jüni, P., Witschi, A., Bloch, R. & Egger, M. (1999). "The hazards of scoring the quality of clinical trials for meta-analysis." *The Journal of the American Medical Association*, 282(11): 1054–1060.

Kabeer, N. (1999). "Resources, Agency, Achievements: Reflections on the Measurement of Women's Empowerment." *Development and Change*, 30: 435–464.

Kabeer, N. (2005). "Is Microfinance a 'Magic Bullet' for Women's Empowerment? Analysis of Findings from South Asia." *Economic and Political Weekly*, October 29, 2005.

Kennedy, C. E., Fonner, V. A., Sweat, M. D., Okero, F. A., Baggaley, R. & O'Reilly, K. R. (2012). "Provider-initiated HIV testing and counseling in low- and middle-income countries: A systematic review." *AIDS Behav.*, Jul 3.

Kennedy, C. E., Medley, A. M., Sweat, M. D. & O'Reilly, K.R. (2010). "Behavioural interventions for HIV positive prevention in developing countries: a systematic review and meta-analysis." *Bull World Health Organ*, 88(8): 615–623.

Kennedy, C. E., Spaulding, A. B., Brickley, D. B., Almers, L., Mirjahangir, J., Packel, L., Kennedy, G.E., Mbizvo, M., Collins, L. & Osborne, K. (2010). "Linking sexual and reproductive health and HIV interventions: a systematic review." *J Int AIDS Soc.*, Jul 19:13:26.

Khodakivska, A. (2013). "Establishing a baseline for lending to women-owned SMEs." International Finance Corporation.

Krishnan, S., Rocca, C., Hubbard, A. E., Subbiah, K., Edmeades, J. & Padian, N.S. (2010). "Do changes in spousal employment status lead to domestic violence? Insights from a prospective study in Bangalore, India." *Social Science & Medicine*, 70: 136–143.

Kundu, A. & Mukherjee, A. K. (2011). "Impact of Swarnajayanti Gram Swarojgar Yojana on health, education and women's empowerment." MPRA Paper No. 33258, Munich Personal RePEc Archive.

Lipsey, M. W. & Wilson, D. B. (2001). *Practical Meta-Analysis*. Thousand Oaks, CA: SAGE.

- Lopez-Claros, A. and Saadia, Z. (2005). "Women's Empowerment: Measuring the Global Gender Gap." World Economic Forum.
- Malhotra, A., Schuler, S.R. & Boender, C. (2002). "Measuring Women's Empowerment as a Variable in International Development." International Center for Research on Women and the Gender and Development Group of the World Bank.
- Mansuri, G. & Rao, V. (2004). "Community-Based and Driven Development: A Critical Review." *World Bank Research Observer*, 19(1): 1–39.
- Mayoux, L. (1998). Women's empowerment and micro-finance programmes: Strategies for increasing impact. *Dev Pract.*, May, 8(2): 235–241.
- Mayoux L. (2001). "Tackling the Down Side: Social Capital, Women's Empowerment and Micro-Finance in Cameroon." *Development and Change*, 23:421-450.
- Medley, A., Kennedy, C., O'Reilly, K., Sweat, M. (2009). "Effectiveness of peer education interventions for HIV prevention in developing countries: a systematic review and meta-analysis." *AIDS Educ Prev.*, Jun., 21(3):181–206.
- Narayan, D. with Patel, R., Schafft, K., Rademacher, A., and Koch-Schulte, S. (2000). *Voices of the Poor: Can Anyone Hear Us?* New York: Oxford University Press.
- Nguyen, V. (2005). "Antiretroviral Globalism, Biopolitics, And Therapeutic Citizenship." *Global Assemblages: Technology, Politics, and Ethics as Anthropological Problems*. Massachusetts: Blackwell Publishing.
- Noblit, G. W. & Hare, R. D., (1988). *Meta-Ethnography: Synthesizing Qualitative Studies*. Newbury Park: Sage.
- Noyes, J., Booth, A., Hannes, K., Harden, A., Harris, J., Lewin, S. & Lockwood, C. (eds.). (2011). *Supplementary Guidance for Inclusion of Qualitative Research in Cochrane Systematic Reviews of Interventions (Version 1, updated August 2011)*. Cochrane Collaboration Qualitative Methods Group.
- Peacock, D. (2013). "South Africa's Sonke Gender Justice Network: Educating men for gender equality." *AGENDA: Empowering women for gender equity*. DOI:10.1080/10130950.2013.808793
- Pritchett, L. & Sandefur, J. (2013). *Context Matters for Size: Why External Validity Claims and Development Practice Don't Mix*. Center for Global Development, Working Paper 336.
- Pronyk, P. M., Hargreaves, J., Kim, J. C., Morison, L. A., Phetla, G., Watts, C., Busza, J. & Porter, J. D. H. (2006). "Effect of a structural intervention for the prevention of intimate-partner violence and HIV in rural South Africa: A cluster randomised trial." *Lancet*, 368: 1973–1983.
- Pulerwitz, J. & Barker, G. (2008). "Measuring attitudes towards gender norms in Brazil: Development and psychometric evaluation of the GEM scale." *Men & Masculinities*, 10: 322–338.
- Roodman, D. (2011). *Due diligence: An impertinent inquiry into microfinance*. Center for Global Development, Washington DC.

SaveAct.org. (2010). "What is the savings and credit group (SCG) model/SaveAct's model?" SaveAct's Model N.p., Web. (Accessed Mar. 6 2013). Available at <http://www.saveact.org.za/about/saveacts-model>.

Sen, A. (2001). "The many faces of gender inequality." *The New Republic*.

A Sinha, PC Parida, P Baurah (2012) The impact of NABARD's Self Help Group-Bank Linkage Programme on poverty and empowerment in India. *Contemporary South Asia*, 2012, 20(4).

Spaulding, A. B, Brickley, D. B., Kennedy, C., Almers, L., Packel, L., Mirjahangir, J., Kennedy, G., Collins, L., Osborne, K., Mbizvo, M. (2009). "Linking family planning with HIV/AIDS interventions: a systematic review of the evidence." *AIDS*. Nov. 23, Suppl 1: S79–88.

Stewart, R., van Rooyen, C. & de Wet, T. (2011). *Do Micro-Credit, Micro-Savings and Micro-Leasing Serve as Effective Financial Inclusion Interventions Enabling Poor People, and Especially Women, to Engage in Meaningful Economic Opportunities in LMICs?* London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Stewart, R., van Rooyen, C., Dickson, K., Majoro, M. & de Wet, T. (2010). *What Is the Impact of Microfinance on Poor People? A Systematic Review of Evidence from Sub-Saharan Africa*. London: EPPI-Centre, Social Science Research Unit, University of London.

Sweat, M. D., Denison, J., Kennedy, C., Tedrow, V. & O'Reilly, K. (2012). "Effects of Condom Social Marketing on Condom Use in Developing Countries: A Systematic Review and Meta-Analysis, 1990-2010." *Bull World Health Organ.*, Aug 1, 90(8): 613–622A.

Thomas, J., Harden, A., Oakley, A., Oliver, S., Sutcliffe, K., Rees, R., Brunton, G. & Kavanagh, J (2004). "Integrating qualitative research with trials in systematic reviews." *BMJ* 2004;328;1010-1012

United Nations. (2000). "Progress of the world's women." New York, NY.

United Nations Development Group (UNDP) (2010). "Thematic Paper On Mdg 3: Promote Gender Equality And Empower Women" March 2010.

United Nations Population Information Network (POPIN) & United Nations Population Fund (UNFPA) (1996). "Guidelines on women's empowerment." Document of the Task Force on ICPD Implementation. New York, NY.

UN Women Watch. (2000). "Five year review for the Beijing Plus five and the platform of action (Beijing Plus 5) held in General Assembly June 5–9, 2000." Retrieved from <http://www.un.org/womenwatch/daw/followup/beijing+5.htm>.

UNESCO. (2013). "International Literacy Data 2013." UNESCO Institute for Statistics.

Upadhyay, U.D. and M.J. Hindin. 2005. Do Higher Status and More Autonomous Women Have Longer Birth Intervals? Results from Cebu, Philippines. *Soc Sci Med* 60(11):2641-2655.

Vaessen, J., Leeuw, F., Bonilla, S., Rivas, A., Lukach, R., Bastiaensen, J., et al. (2014). *The effect of microcredit on women's control over household spending in developing countries*. The Campbell Library, March 2014.

Van Kempen, L. (2009). "The 'Downside' of Women's empowerment in India: An Experimental Inquiry into the Role of Expectations." *Social Indicators Research*, 94(3): 465–482.

Waddington, H., White, H., Snilstveit, B., Hombrados, J. G., Vojtkova, M., Davies, P., Bhavsar, A., Eysers, J., Koehlmoos, T. P., Petticrew, M., Valentine, J. C. & Tugwell, P. (2012). "How to do a good systematic review of effects in international development: A tool kit." *Journal of Development Effectiveness*, 4(3): 359–387.

Waddington, H., Snilstveit, B., Hombrados, J., Vojtkova, M., Phillips, D., Davies, P. & White, H. (Forthcoming). "Farmer Field Schools for Improving Farming Practices and Farmer Outcomes: A Systematic Review." International Initiative for Impact Evaluation.

Walsh, D., & Downe, S. (2005) "Meta-synthesis method for qualitative research: a literature review." *Journal of Advanced Nursing*, 50(2): 204–211.

White, H. (2014). "Current Challenges in Impact Evaluation" *European Journal of Development Research*, 26: 18–30.

White, H., & Waddington, H. (2012). "Why do we care about evidence synthesis? An introduction to the special issue on systematic reviews." *Journal of Development Effectiveness*, 4(3), 351–358.

White, H., & Phillips, D. (2012). Addressing attribution of cause and effect in small n impact evaluations: Towards an integrated framework (3ie Working Paper 15). New Delhi, India: 3ie.

Woolcock, M. (2013). "Using case studies to explore the external validity of 'complex' development interventions." *Evaluation*, 19(3): 229–248.

World Bank. (2009). "Building on Tradition as the way to Women's Empowerment in Cambodia." *East Asia and Pacific Region Social Development Notes*.

World Health Organization (WHO) (2007). "Unequal, Unfair, Ineffective And Inefficient Gender Inequity In Health: Why It Exists And How We Can Change It." Final Report to the WHO Commission on Social Determinants of Health, Geneva, Switzerland.

Appendix 1: Data extraction form

Study Data Extraction/Coding

Study ID (sid):

Coders Initials (coderid):

Date Coded (date):

Author(s) (author):

Funder (funder):

Publication date (pubdate):

Country (country):

Start date of study (startdate):

Start date of study (enddate):

Publication type (pubtype): (1) Book, (2) peer-reviewed journal, (3) book chapter, (4) dissertation/thesis, (5) unpublished report

SHG Data Extraction/Coding

Study ID (sid):

Coders initials (coderid):

Date coded (date):

Name of self-help group (shgname):

Location of group (glocale):

Region (gregion):

Target population (targetpop):

Type of group (gtype): (1) economic, (2) livelihood, (3) other

Number of intervention components (numcomp):

Type of component: (1) credit, (2) savings, (3) loans, (4) insurance (5) capacity building:

Type of component 1 (comp1)

Type of component 2 (comp2)

Type of component 3 (comp3)

Type of component 4 (comp4)

Type of component 5 (comp5)

Group origin (origin): (1) community-based, (2) organization-based (3) research-based

Study design (design):

Nature of comparison group (compgroup):

Sample size (sampsize):

Type of sampling (samptype): (1) random (2) purposive (3) convenience (4) cannot tell

Did researchers assess baseline differences? (basediff)

If yes, were there differences? (difftype) (1) no (2) minor (3) major (4) cannot tell

Outcome Extraction/Coding

Study ID (sid):

Coders initials (coderid):

Date coded (date):

Outcome category (outcat): (1) economic (2) political (3) social (4) psychological

Outcome name (outname):

Type of information (outtype): (1) quantitative (2) qualitative

Source of information (outsource): (1) survey (2) records (3) interviews (4) focus groups

Measure/Indicator of outcome (measure):

Were there any differences in measurement of this outcome between the group participants and the comparison? (1) yes (2) no (3) cannot tell

Effect Size Extraction/Coding

Study ID (sid):

Coders initials (coderid):

Date coded (date):

Outcome category (outcat): (1) economic (2) political (3) social (4) psychological

Outcome name (outname):

Direction of effect (esdir): (1) effect favors self-help group (2) effect favors comparison (3) effect favors neither (4) cannot tell

Effect is statistically significant (essig)? (1) yes (2) no (3) cannot tell

SHG sample size (shgss):

Comparison sample size (compss):

For continuous measures:

SHG group mean (txmean):

Comparison group mean (compmean):

Are means reported above adjusted? (meanadj): (1) yes (2) no

SHG group standard deviation (txsd):

Comparison group standard deviation (compsd):

SHG group standard error (txse):

Comparison group standard error (compse):

t-value from an independent *t*-test (est)

For dichotomous measures:

SHG group number of participants who experienced a change (txnum):

Comparison group number of participants who experienced a change (comprnum):

SHG group proportion of participants who experienced a change (txpro):

Comparison group proportion of participants who experienced a change (comppro):

Are the proportions above adjusted for pretest variables? (proadj): (1) yes (2) no

Logged odds-ratio (eslgodd):

Standard error of logged odds-ratio (eslgoddse):

Logged odds-ratio adjusted? (e.g., from a logistic regression analysis with other independent variables) (1=yes; 0=no)

Chi-square value with df = 1 (2 by 2 contingency table) (eschi):

Correlation coefficient (esphi):

For Hand Calculated Data:

Hand calculated d-type effect size (eshand1)

Hand calculated standard error of the d-type effect size (eshand2)

Hand calculated odds-ratio effect size (eshand3)

Hand calculated odds-ratio standard error (eshand4)

Intermediate outcomes or themes (knowledge, skills):

For qualitative data:

Participants views (views):

Themes (mtheme):

Subthemes (sttheme):

Sources: Wilson et al.

Appendix 2: Full search strategy

Search	Query	Items found
#5	Search ((#1) AND #2) AND #3 Filters: Publication date from 1980/01/01 to 2012/12/31	1741
#4	Search ((#1) AND #2) AND #3	1811
#3	Search “women’s self-help”[tiab] OR “women’s cooperative*”[tiab] OR “self-help group*”[tiab] OR “self help group*”[tiab] OR “support group*”[tiab] OR “lending group*”[tiab] OR “advocacy group*”[tiab] OR “micro finance”[tiab] OR “micro credit”[tiab] OR “microfinance”[tiab] OR “microcredit”[tiab] OR “income generation group*”[tiab] OR “microenterprise group*”[tiab] OR sangha[tiab] OR “Self-Help Groups”[Mesh] OR (women*[tiab] AND (financ*[tiab] OR economic*[tiab])) OR (“Women”[Mesh] AND (“Financing, Organized”[Mesh] OR “Economics”[Mesh]))	29946
#2	Search “women’s empowerment”[tiab] OR “empower*” [tiab] OR “girl’s empowerment”[tiab] OR “empowering”[tiab] OR “power”[tiab] OR “control”[tiab] OR “Power (Psychology)”[Mesh]	1743835
#1	Search (“developing country”[tiab] OR “developing countries”[tiab] OR “developing nation”[tiab] OR “developing nations”[tiab] OR “developing population”[tiab] OR “developing populations”[tiab] OR “developing world”[tiab] OR “less developed country”[tiab] OR “less developed countries”[tiab] OR “less developed nation”[tiab] OR “less developed nations”[tiab] OR “less developed population”[tiab] OR “less developed populations”[tiab] OR “less developed world”[tiab] OR “lesser developed country”[tiab] OR “lesser developed countries”[tiab] OR “lesser developed nation”[tiab] OR “lesser developed nations”[tiab] OR “lesser developed population”[tiab] OR “lesser developed populations”[tiab] OR “lesser developed world”[tiab] OR “under developed country”[tiab] OR “under developed countries”[tiab] OR “under developed nation”[tiab] OR “under developed nations”[tiab] OR “under developed population”[tiab] OR “under developed populations”[tiab] OR “under developed world”[tiab] OR “underdeveloped country”[tiab] OR “underdeveloped countries”[tiab] OR “underdeveloped nation”[tiab] OR “underdeveloped nations”[tiab] OR “underdeveloped population”[tiab] OR “underdeveloped populations”[tiab] OR “underdeveloped world”[tiab] OR “middle income country”[tiab] OR “middle income countries”[tiab] OR “middle income nation”[tiab] OR “middle income nations”[tiab] OR “middle income population”[tiab] OR “middle income populations”[tiab] OR “low income country”[tiab] OR “low income countries”[tiab] OR “low income nation”[tiab] OR “low income nations”[tiab] OR “low income population”[tiab] OR “low income populations”[tiab] OR “lower income country”[tiab] OR “lower income countries”[tiab] OR “lower income nation”[tiab] OR “lower income nations”[tiab]	1139069

	OR "lower income population"[tiab] OR "lower income populations"[tiab] OR "underserved country"[tiab] OR "underserved countries"[tiab] OR "underserved nation"[tiab] OR "underserved nations"[tiab] OR "underserved population"[tiab] OR "underserved populations"[tiab] OR "underserved world"[tiab] OR "under served country"[tiab] OR "under served countries"[tiab] OR "under served nation"[tiab] OR "under served nations"[tiab] OR "under served population"[tiab] OR "under served populations"[tiab] OR "under served world"[tiab] OR "deprived country"[tiab] OR "deprived countries"[tiab] OR "deprived nation"[tiab] OR "deprived nations"[tiab] OR "deprived population"[tiab] OR "deprived populations"[tiab] OR "deprived world"[tiab] OR "poor country"[tiab] OR "poor countries"[tiab] OR "poor nation"[tiab] OR "poor nations"[tiab] OR "poor population"[tiab] OR "poor populations"[tiab] OR "poor world"[tiab] OR "poorer country"[tiab] OR "poorer countries"[tiab] OR "poorer nation"[tiab] OR "poorer nations"[tiab] OR "poorer population"[tiab] OR "poorer populations"[tiab] OR "poorer world"[tiab] OR "developing economy"[tiab] OR "developing economies"[tiab] OR "less developed economy"[tiab] OR "less developed economies"[tiab] OR "lesser developed economy"[tiab] OR "lesser developed economies"[tiab] OR "under developed economy"[tiab] OR "under developed economies"[tiab] OR "underdeveloped economy"[tiab] OR "underdeveloped economies"[tiab] OR "middle income economy"[tiab] OR "middle income economies"[tiab] OR "low income economy"[tiab] OR "low income economies"[tiab] OR "lower income economy"[tiab] OR "lower income economies"[tiab] OR "low gdp"[tiab] OR "low gnp"[tiab] OR "low gross domestic"[tiab] OR "low gross national"[tiab] OR "lower gdp"[tiab] OR "lower gnp"[tiab] OR "lower gross domestic"[tiab] OR "lower gross national"[tiab] OR lmic[tiab] OR Imics[tiab] OR "third world"[tiab] OR "lami country"[tiab] OR "lami countries"[tiab] OR "transitional country"[tiab] OR "transitional countries"[tiab] OR "resource-limited"[tiab] OR "resource-constrained"[tiab]) OR (Africa[tiab] OR Asia[tiab] OR Caribbean[tiab] OR West Indies[tiab] OR South America[tiab] OR Latin America[tiab] OR Central America[tiab] OR Afghanistan[tiab] OR Albania[tiab] OR Algeria[tiab] OR Angola[tiab] OR Antigua[tiab] OR Barbuda[tiab] OR Argentina[tiab] OR Armenia[tiab] OR Armenian[tiab] OR Aruba[tiab] OR Azerbaijan[tiab] OR Bahrain[tiab] OR Bangladesh[tiab] OR Barbados[tiab] OR Benin[tiab] OR Byelarus[tiab] OR Byelorussian[tiab] OR Belarus[tiab] OR Belorussian[tiab] OR Belorussia[tiab] OR Belize[tiab] OR Bhutan[tiab] OR Bolivia[tiab] OR Bosnia[tiab] OR Herzegovina[tiab] OR Hercegovina[tiab] OR Botswana[tiab] OR Brasil[tiab] OR Brazil[tiab] OR Bulgaria[tiab] OR Burkina Faso[tiab] OR Burkina Fasso[tiab] OR Upper Volta[tiab] OR Burundi[tiab] OR Urundi[tiab] OR Cambodia[tiab] OR Khmer Republic[tiab] OR Kampuchea[tiab] OR Cameroon[tiab] OR Cameroons[tiab] OR Cameron[tiab] OR Cape Verde[tiab] OR Central African Republic[tiab] OR Chad[tiab] OR Chile[tiab] OR China[tiab] OR Colombia[tiab] OR Comoros[tiab] OR Comoro Islands[tiab] OR Comores[tiab] OR Mayotte[tiab] OR Congo[tiab] OR Zaire[tiab] OR Costa Rica[tiab] OR Cote d'Ivoire[tiab] OR Ivory Coast[tiab] OR Croatia[tiab] OR Cuba[tiab] OR Cyprus[tiab] OR Czechoslovakia[tiab] OR Czech Republic[tiab] OR Slovakia[tiab] OR Slovak Republic[tiab] OR Djibouti[tiab] OR French	
--	---	--

	Somaliland[tiab] OR Dominica[tiab] OR Dominican Republic[tiab] OR East Timor[tiab] OR Timor Leste[tiab] OR Ecuador[tiab] OR Egypt[tiab] OR United Arab Republic[tiab] OR El Salvador[tiab] OR Eritrea[tiab] OR Estonia[tiab] OR Ethiopia[tiab] OR Fiji[tiab] OR Gabon[tiab] OR Gabonese Republic[tiab] OR Gambia[tiab] OR Gaza[tiab] OR Georgia Republic[tiab] OR Georgian Republic[tiab] OR Ghana[tiab] OR Gold Coast[tiab] OR Greece[tiab] OR Grenada[tiab] OR Guatemala[tiab] OR Guinea[tiab] OR Guam[tiab] OR Guiana[tiab] OR Guyana[tiab] OR Haiti[tiab] OR Honduras[tiab] OR Hungary[tiab] OR India[tiab] OR Maldives[tiab] OR Indonesia[tiab] OR Iran[tiab] OR Iraq[tiab] OR Isle of Man[tiab] OR Jamaica[tiab] OR Jordan[tiab] OR Kazakhstan[tiab] OR Kazakh[tiab] OR Kenya[tiab] OR Kiribati[tiab] OR Korea[tiab] OR Kosovo[tiab] OR Kyrgyzstan[tiab] OR Kirghizia[tiab] OR Kyrgyz Republic[tiab] OR Kirghiz[tiab] OR Kirgizstan[tiab] OR "Lao PDR"[tiab] OR Laos[tiab] OR Latvia[tiab] OR Lebanon[tiab] OR Lesotho[tiab] OR Basutoland[tiab] OR Liberia[tiab] OR Libya[tiab] OR Lithuania[tiab] OR Macedonia[tiab] OR Madagascar[tiab] OR Malagasy Republic[tiab] OR Malaysia[tiab] OR Malaya[tiab] OR Malay[tiab] OR Sabah[tiab] OR Sarawak[tiab] OR Malawi[tiab] OR Nyasaland[tiab] OR Mali[tiab] OR Malta[tiab] OR Marshall Islands[tiab] OR Mauritania[tiab] OR Mauritius[tiab] OR Mexico[tiab] OR Micronesia[tiab] OR Middle East[tiab] OR Moldova[tiab] OR Moldovia[tiab] OR Moldovian[tiab] OR Mongolia[tiab] OR Montenegro[tiab] OR Morocco[tiab] OR Ifni[tiab] OR Mozambique[tiab] OR Myanmar[tiab] OR Myanma[tiab] OR Burma[tiab] OR Namibia[tiab] OR Nepal[tiab] OR Netherlands Antilles[tiab] OR New Caledonia[tiab] OR Nicaragua[tiab] OR Niger[tiab] OR Nigeria[tiab] OR Northern Mariana Islands[tiab] OR Oman[tiab] OR Muscat[tiab] OR Pakistan[tiab] OR Palau[tiab] OR Palestine[tiab] OR Panama[tiab] OR Paraguay[tiab] OR Peru[tiab] OR Philippines[tiab] OR Philipines[tiab] OR Phillipines[tiab] OR Phillippines[tiab] OR Poland[tiab] OR Portugal[tiab] OR Puerto Rico[tiab] OR Romania[tiab] OR Rumania[tiab] OR Roumania[tiab] OR Russia[tiab] OR Russian[tiab] OR Rwanda[tiab] OR Ruanda[tiab] OR Saint Kitts[tiab] OR St Kitts[tiab] OR Nevis[tiab] OR Saint Lucia[tiab] OR St Lucia[tiab] OR Saint Vincent[tiab] OR St Vincent[tiab] OR Grenadines[tiab] OR Samoa[tiab] OR Samoan Islands[tiab] OR Sao Tome[tiab] OR Saudi Arabia[tiab] OR Senegal[tiab] OR Serbia[tiab] OR Montenegro[tiab] OR Seychelles[tiab] OR Sierra Leone[tiab] OR Slovenia[tiab] OR Sri Lanka[tiab] OR Ceylon[tiab] OR Solomon Islands[tiab] OR Somalia[tiab] OR Sudan[tiab] OR Suriname[tiab] OR Surinam[tiab] OR Swaziland[tiab] OR Syria[tiab] OR Tajikistan[tiab] OR Tadzhikistan[tiab] OR Tadjikistan[tiab] OR Tadjhik[tiab] OR Tanzania[tiab] OR Thailand[tiab] OR Togo[tiab] OR Togolese Republic[tiab] OR Tonga[tiab] OR Trinidad[tiab] OR Tobago[tiab] OR Tunisia[tiab] OR Turkey[tiab] OR Turkmenistan[tiab] OR Turkmen[tiab] OR Uganda[tiab] OR Ukraine[tiab] OR Uruguay[tiab] OR USSR[tiab] OR Soviet Union[tiab] OR Union of Soviet Socialist Republics[tiab] OR Uzbekistan[tiab] OR Uzbek OR Vanuatu[tiab] OR New Hebrides[tiab] OR Venezuela[tiab] OR Vietnam[tiab] OR Viet Nam[tiab] OR West Bank[tiab] OR Yemen[tiab] OR Yugoslavia[tiab] OR Zambia[tiab] OR Zimbabwe[tiab]) OR (Developing Countries[Mesh:noexp] OR Africa[Mesh:noexp] OR Africa,	
--	---	--

	Northern[Mesh:noexp] OR Africa South of the Sahara[Mesh:noexp] OR Africa, Central[Mesh:noexp] OR Africa, Eastern[Mesh:noexp] OR Africa, Southern[Mesh:noexp] OR Africa, Western[Mesh:noexp] OR Asia[Mesh:noexp] OR Asia, Central[Mesh:noexp] OR Asia, Southeastern[Mesh:noexp] OR Asia, Western[Mesh:noexp] OR Caribbean Region[Mesh:noexp] OR West Indies[Mesh:noexp] OR South America[Mesh:noexp] OR Latin America[Mesh:noexp] OR Central America[Mesh:noexp] OR Afghanistan[Mesh:noexp] OR Albania[Mesh:noexp] OR Algeria[Mesh:noexp] OR American Samoa[Mesh:noexp] OR Angola[Mesh:noexp] OR "Antigua and Barbuda"[Mesh:noexp] OR Argentina[Mesh:noexp] OR Armenia[Mesh:noexp] OR Azerbaijan[Mesh:noexp] OR Bahrain[Mesh:noexp] OR Bangladesh[Mesh:noexp] OR Barbados[Mesh:noexp] OR Benin[Mesh:noexp] OR Byelarus[Mesh:noexp] OR Belize[Mesh:noexp] OR Bhutan[Mesh:noexp] OR Bolivia[Mesh:noexp] OR Bosnia-Herzegovina[Mesh:noexp] OR Botswana[Mesh:noexp] OR Brazil[Mesh:noexp] OR Bulgaria[Mesh:noexp] OR Burkina Faso[Mesh:noexp] OR Burundi[Mesh:noexp] OR Cambodia[Mesh:noexp] OR Cameroon[Mesh:noexp] OR Cape Verde[Mesh:noexp] OR Central African Republic[Mesh:noexp] OR Chad[Mesh:noexp] OR Chile[Mesh:noexp] OR China[Mesh:noexp] OR Colombia[Mesh:noexp] OR Comoros[Mesh:noexp] OR Congo[Mesh:noexp] OR Costa Rica[Mesh:noexp] OR Cote d'Ivoire[Mesh:noexp] OR Croatia[Mesh:noexp] OR Cuba[Mesh:noexp] OR Cyprus[Mesh:noexp] OR Czechoslovakia[Mesh:noexp] OR Czech Republic[Mesh:noexp] OR Slovakia[Mesh:noexp] OR Djibouti[Mesh:noexp] OR "Democratic Republic of the Congo"[Mesh:noexp] OR Dominica[Mesh:noexp] OR Dominican Republic[Mesh:noexp] OR East Timor[Mesh:noexp] OR Ecuador[Mesh:noexp] OR Egypt[Mesh:noexp] OR El Salvador[Mesh:noexp] OR Eritrea[Mesh:noexp] OR Estonia[Mesh:noexp] OR Ethiopia[Mesh:noexp] OR Fiji[Mesh:noexp] OR Gabon[Mesh:noexp] OR Gambia[Mesh:noexp] OR "Georgia (Republic)"[Mesh:noexp] OR Ghana[Mesh:noexp] OR Greece[Mesh:noexp] OR Grenada[Mesh:noexp] OR Guatemala[Mesh:noexp] OR Guinea[Mesh:noexp] OR Guinea-Bissau[Mesh:noexp] OR Guam[Mesh:noexp] OR Guyana[Mesh:noexp] OR Haiti[Mesh:noexp] OR Honduras[Mesh:noexp] OR Hungary[Mesh:noexp] OR India[Mesh:noexp] OR Indonesia[Mesh:noexp] OR Iran[Mesh:noexp] OR Iraq[Mesh:noexp] OR Jamaica[Mesh:noexp] OR Jordan[Mesh:noexp] OR Kazakhstan[Mesh:noexp] OR Kenya[Mesh:noexp] OR Korea[Mesh:noexp] OR Kosovo[Mesh:noexp] OR Kyrgyzstan[Mesh:noexp] OR Laos[Mesh:noexp] OR Latvia[Mesh:noexp] OR Lebanon[Mesh:noexp] OR Lesotho[Mesh:noexp] OR Liberia[Mesh:noexp] OR Libya[Mesh:noexp] OR Lithuania[Mesh:noexp] OR Macedonia[Mesh:noexp] OR Madagascar[Mesh:noexp] OR Malaysia[Mesh:noexp] OR Malawi[Mesh:noexp] OR Mali[Mesh:noexp] OR Malta[Mesh:noexp] OR Mauritania[Mesh:noexp] OR Mauritius[Mesh:noexp] OR Mexico[Mesh:noexp] OR Micronesia[Mesh:noexp] OR Middle East[Mesh:noexp] OR Moldova[Mesh:noexp] OR Mongolia[Mesh:noexp] OR Montenegro[Mesh:noexp] OR Morocco[Mesh:noexp] OR	
--	---	--

	Mozambique[Mesh:noexp] OR Myanmar[Mesh:noexp] OR Namibia[Mesh:noexp] OR Nepal[Mesh:noexp] OR Netherlands Antilles[Mesh:noexp] OR New Caledonia[Mesh:noexp] OR Nicaragua[Mesh:noexp] OR Niger[Mesh:noexp] OR Nigeria[Mesh:noexp] OR Oman[Mesh:noexp] OR Pakistan[Mesh:noexp] OR Palau[Mesh:noexp] OR Panama[Mesh:noexp] OR Papua New Guinea[Mesh:noexp] OR Paraguay[Mesh:noexp] OR Peru[Mesh:noexp] OR Philippines[Mesh:noexp] OR Poland[Mesh:noexp] OR Portugal[Mesh:noexp] OR Puerto Rico[Mesh:noexp] OR Romania[Mesh:noexp] OR Russia[Mesh:noexp] OR Rwanda[Mesh:noexp] OR "Saint Kitts and Nevis"[Mesh:noexp] OR Saint Lucia[Mesh:noexp] OR "Saint Vincent and the Grenadines"[Mesh:noexp] OR Samoa[Mesh:noexp] OR Saudi Arabia[Mesh:noexp] OR Senegal[Mesh:noexp] OR Serbia[Mesh:noexp] OR Montenegro[Mesh:noexp] OR Seychelles[Mesh:noexp] OR Sierra Leone[Mesh:noexp] OR Slovenia[Mesh:noexp] OR Sri Lanka[Mesh:noexp] OR Somalia[Mesh:noexp] OR South Africa[Mesh:noexp] OR Sudan[Mesh:noexp] OR Suriname[Mesh:noexp] OR Swaziland[Mesh:noexp] OR Syria[Mesh:noexp] OR Tajikistan[Mesh:noexp] OR Tanzania[Mesh:noexp] OR Thailand[Mesh:noexp] OR Togo[Mesh:noexp] OR Tonga[Mesh:noexp] OR "Trinidad and Tobago"[Mesh:noexp] OR Tunisia[Mesh:noexp] OR Turkey[Mesh:noexp] OR Turkmenistan[Mesh:noexp] OR Uganda[Mesh:noexp] OR Ukraine[Mesh:noexp] OR Uruguay[Mesh:noexp] OR USSR[Mesh:noexp] OR Uzbekistan[Mesh:noexp] OR Vanuatu[Mesh:noexp] OR Venezuela[Mesh:noexp] OR Vietnam[Mesh:noexp] OR Yemen[Mesh:noexp] OR Yugoslavia[Mesh:noexp] OR Zaire[Mesh:noexp] OR Zambia[Mesh:noexp] OR Zimbabwe[Mesh:noexp])	
--	--	--

Appendix 3: Search diary

Databases

Using PUBMED Search Strategy

9-3-2013: **PUB MED**: Total hits 1320

9-3-2013: **POPLINE**: Total hits 86

9-3-2013: **PYSCHMED**: total hits 67

3-20-2013 **3ie Database**: 17 articles found; 4 included

3-8-2013 **JOLIS : IMF**: 392 articles; none included

3-9-2013 **JOLIS: World Bank**: 1239 Results; 11 included

Using Alternative Search Strategies:

3-1-2013 **PROquest Social Sciences**: women OR woman OR Female OR Girl OR Self-help OR self help OR support OR empower OR women's empowerment OR girl's empowerment OR empowering OR power OR control OR decision-making OR choice OR violence OR cooperative OR collective OR program OR Group OR organization, 2046 found; 0 included

4-3-2012: **IBSS International Bibliography of Social Sciences/Proquest**, 431 results, 7 included

INDMED: Searched: group AND economic; women AND group; woman OR women OR female OR girl AND group OR cooperative OR program OR collective AND empowerment OR empower OR empowering OR power OR control OR choice OR violence: 0 found, 0 included

Search: empowerment: Total HITS: 18, Included: 6

12-31-2012 **Index Medicus for the WHO** <http://www.globalhealthlibrary.net> : Search: woman OR women OR female OR girl AND group OR cooperative OR program OR collective AND empowerment OR empower OR empowering OR power OR control OR choice OR violence: 29 results; 0 included

Search: empowerment AND group AND women, 207 results; Included: 14

BLDS: Search: woman OR women OR female OR girl AND group OR cooperative OR program OR collective AND empowerment OR empower OR empowering OR power OR control OR choice OR violence: No results

Search: "empowerment AND group" : 15 results; Included 7

Search: "women AND empower AND group" returned 3 results – already included relevant results

Search: "women AND empowerment AND group " returned 10 results – already included relevant results

Search: "women AND group" returned 219 results, Included: none

AFRICABIB <http://www.africabib.org/>: Search: empowerment AND group: Found: 17, Included: 1; Search: empowerment : Found: 449, Included: 2

African Women Bibliographic Database, Search: empowerment AND group: Found: 4, Included: 0; Search: empowerment : Found: 439, Included: 2

Women Travelers Bibliographic Database: Search: empowerment AND group: Found: 0, Included: 0; Search: empowerment : Found: 0, Included: 0

Islam in Contemporary Sub-Saharan Africa, Search: empowerment: Found: 5, Included: 0

Kenya Coast Bibliographic Database, Search: empowerment: Found: 0

EconLit, Search: empowerment : Found: 4, included: 0

FEMNET: Hits: 3, Included: 0

Multilateral Organizations

Keywords Search: Women AND Group AND Empower*

10-18-13: WHO website, 24 articles searched; 3 include; 5 maybe

10-15-13: USAID, 9 articles searched; 2 maybe

10-14-13: United Nations Development Fund, 13 articles searched; 2 maybe studies; 0 include

10-12-13: United Kingdom of International Development, 25 articles searched; 6 studies included; 1 maybe

7-6-2013: Journal of International Development, 10 articles searched; 4 included

6-7-2013: African Development Bank, 24 articles searched; 3 included; 3 Maybe

3-29-2012 Journal: Economic development and cultural change, 2005-2013: 5 included

12-27-2012 Google Scholar; 8 included

12-28-2012 Asian Development Bank; 3 included,

African Development Bank; 5 included

UNICEF; 3 included,

United Nations Development Programme; 4 included,

12-17-2012: United Nations Fund for Population; 4 included

United Nations Development Fund For Women; 9 included,

12-31-2012, UNDP, 3 included

12-28-2012, African Development Bank, 4 included

UNICEF; 3 included

12-27-2012, United Nations Development Programme; 4 included

12-17-2012, United Nations Fund for Population; 4 included

United Nations Development Fund For Women, 9 included

Inter-American Development Bank, 2 included

International Food Policy Research Institute, 1 included

12/21/2012, United Kingdom's Department for International Development ; 13 included

United States Agency for International Development ; 14 included

World Bank; 3 included

International Fund for Agricultural Development; 2 included

World Health Organization; 4 included

Inter-American Development Bank; 2 included

International Food Policy Research Institute; 1 included

Hand Search of Websites

SEWA, 1 included

AED Center for Gender Equality, 0 Included

Asian Women's Network on Gender and Development, 0 Included

The Center for the Evaluation of Global Action, 0 Included

Ford Foundation, 0 Included

Global Fund for Women, 0 Included

GROOTS International, 1 included

The Guttmacher Institute, 0 Included

The Hewlett Foundation, 0 Included

International Committee for Research on Women, 3 included

Latin American Women and Habitat Network, 0 Included

The Packard Foundation, 0 Included
UCGHI Center of Expertise on Women's Health and Empowerment, 0 Included
Women Deliver, 1 included

Journal Search in Library

12-12-12: Health Policy, 0 included
Global Public Health, 0 included
Indian Journal of Gender Studies: Total hits: 606; Included: 20
07-11-13: Third World Quarterly: Total hits: 793, Included: 6
7-17-13: Development & Change total hits: 732: Included: 13
Health Care for Women International; 3 included
Development; 15 included
Journal of Development Economics, 0 included
International Journal of Sustainable Development, 0 included
Indian Growth and Development Review, 1 included
Journal of International Development , 9 included
07-11-13: Third World Quarterly, Total hits: 354, Included: 6
Current Anthropology, 4 hits, 0 included
Economic Development and Cultural Change, 3 hits 0 included
Feminist Economics, 85 hits, 1 included
Indian Journal of Gender Studies, 0 included
International Journal of Health Planning and Management, 13 hits, 0 included
International Journal of Social Research Methodology, 0 included
Qualitative Research in Psychology, 0 included
World Development, 0 included

Key Word Search

2-21-2013: Search: Mexico Indigenous People's Development Project; 1 included
Search: Professional Assistance for Development Action; 1 included
Search: Self Employed Womens Association; 1 included
Search: Swarnjayanthi Gram SwarozgarYojana; 3 included

1-24-2013: Search: Livelihood Empowerment Against Poverty; 3 included
Search: Productive Safety Net Programme self-help group women's empowerment; 1 included
Search: Comprehensive Africa Agricultural Development Programme; 1 included
Search: Colombia Humanitarian women's empowerment economic self-help group; 1 included
Search Google Scholar 2 pages: Grassroots Women Environmental Protection and Poverty Alleviation Project; 2 included

1-16-2013: Search: Division of Advancement for Women; 3 included
Search: Gender Equity Model Egypt; 1 included
Search: Redcamif; 0 included
Search: Promujer; 2 included

1-15-2013: Key Contact added: Ushma Uppaday; 1 included

1-10-2013: Search: Progresa; 3 included

Search: The Self-Employed Women's Association (SEWA); 1 included

Search: The Social Entrepreneurship Program; 1 included

Search: Inter-American Center for Research and Documentation on Professional Training; 0 included

Search: Womens World Banking; 0 included

Search: OAS; 0 included

Search: Grameen Bank india micro finance; 5 included

1-5-2013: Bangladesh Rural Advancement Committee (BRAC) India Micro finance; 7 included

1-10-2013: Progressa; 2 included

Search: The Self-Employed Women's Association (SEWA); 1 included

Search: The Social Entrepreneurship Program; 1 included

Search: Inter-American Center for Research and Documentation on Professional Training; 0 included

Search: OAS; 0 included

Search: Grameen Bank india micro finance; 5 included

1-5-2013; Search: Bangladesh Rural Advancement Committee (BRAC) India Micro finance; 7 included

Search: the GSMA mWomen Programme, 1 included

Search: Hauirou Commission, 1 included

Search: The Peri-Urban Interface, 1 included

Appendix 4: Reasons for exclusion of marginal studies

Quantitative	Reason for Exclusion
Ackerly, 1995	This study does not have a valid comparison group.
Ashburn, 2007	This study does not focus on empowerment outcomes.
Banerjee, 2004	No quantitative estimate of impact
Bushamuka et al, 2005	This is not an evaluation of a self-help group program
Chandra and Sinha, 2010	This study does not have a valid comparison group and outcomes are not measured at the woman level
Deininger and Liu, 2009	No focus on empowerment
Euser et al., 2012	There is no clear comparison group
Feigenberg, 2010	No focus on empowerment and this is an RCT but the control group also consists of members of SHGs
Lokhande, 2013	There is no clear comparison group
Madheswaran, 2001	No quantitative estimate of impact
Mansuri, 2010	There is no clear comparison group
Mayoux, 2005	No quantitative estimate of impact
Murthy 2012	This is not an evaluation of a self-help group program
Odutolu et al., 2003	No quantitative estimate of impact
Oosterhoff et al., 2008	No quantitative estimate of impact
Panda, 2009	This study does not focus on empowerment outcomes.
Parajuli, 2012	This is not an evaluation of a self-help group program
Premaratne et al., 2012	There is no clear comparison group
Pronyk et al., 2008	This study does not focus on empowerment outcomes.
Pucho, 2008	This study does not have a valid comparison group.
Reddy, 2005	No quantitative estimate of impact
Sabhlok, 2011	No quantitative estimate of impact
Salway, 2005	This is not an evaluation of a self-help group program
Sinha, Parida and Baurah, 2012	This study does not have a valid comparison group.
Sinha, Pastakia, 2004	No quantitative estimate of impact
Suguna, 2006	This study does not have a valid comparison group.

Swain and Varghese, 2009	This study does not focus on empowerment outcomes.
Teshome et al., 2006	No quantitative estimate of impact
UNFPA, 2006	No quantitative estimate of impact
Urquieta et al., 2009	This is not an evaluation of a self-help group program
Uys, Benghu, and Majumdar, 2006	No quantitative estimate of impact
Van Kempen, 2009	This study does not have a valid comparison group.
Vijayanthi, 2002	No quantitative estimate of impact

<u>Qualitative</u>	Reason
Ahmed, 2011	This is not an evaluation of a self-help group program
Apantaku, 2008	This is not an evaluation of a self-help group program
Barry, 2012	This study does not focus on empowerment outcomes.
Bhat, 2001	This is not an evaluation of a self-help group program
Bhengu, 2010	This is not an evaluation of a self-help group program
Biradavolu, 2009	This study does not report direct quotations from participants
Cheston, 2002	This is not an evaluation of a self-help group program
Faraizi, 2011	This study does not report direct quotations from participants
Ghadoliya, 2003	This is not an evaluation of a self-help group program
Gibb, 2008	This is not an evaluation of a self-help group program
Guerin, 2006	This study does not report direct quotations from participants
Hoodfar, 2010	This is not an evaluation of a self-help group program
Islam, 2011	This is not an evaluation of a self-help group program
Jerinabi, 2008	This study does not report direct quotations from participants
Kim, 2007	This is not an evaluation of a self-help group program
Kuttab, 2010	This is not an evaluation of a self-help group program
Lokhande 2008	This is not an evaluation of a self-help group program
Lombe, 2012	This study does not report direct quotations from women participants
Mayoux, 2000	This is not an evaluation of a self-help group program
Mayoux, 2001	This study does not report direct quotations from participants

Meena, 2011	This is not an evaluation of a self-help group program
Moyle 2006	This study does not report direct quotations from participants
Ndinda, 2009	This is not an evaluation of a self-help group program
Nguyen, 2009	This is not an evaluation of an self-help group program
Nkosi, 2003	This study does not report direct quotations from participants
Noreen, 2011	This is not an evaluation of a self-help group program
Norwood, 2004	This study does not report direct quotations from participants
Panwar, 2010	This is not an evaluation of a self-help group program
Patel, 2010	This study does not report direct quotations from participants
Pronyk, 2008a	This is not an evaluation of a self-help group program
Pronyk, 2008b	This is not an evaluation of a self-help group program
Rahman, 2011	This is not an evaluation of a self-help group program
Reza-Paul, 2012	This is not an evaluation of a self-help group program
Roger, 2011	This is not an evaluation of a self-help group program
Sabhlok, 2011	This study does not focus on empowerment outcomes.
Salway 2005	This is not an evaluation of a self-help group program
Sarojani, 2009	This study does not report direct quotations from participants
Sharma, 2014	This is not an evaluation of a self-help group program
Shylendra, 1999	This study does not report direct quotations from participants
Somé, 2013	This is not an evaluation of a self-help group program
Sotshongaye, 2000	This is not an evaluation of a self-help group program
Ssewamala, 2009	This is not an evaluation of a self-help group program
Torri, 2011	This is not an evaluation of a self-help group program
Tupe, 2013	This study does not report direct quotations from participants
Vijayanthi, 2002	This study does not report direct quotations from participants

Appendix 5: Qualitative study quality assessment

Screening Question: Is there a clear statement of study aims of the research?	Yes / Can't tell / No
Screening Question: Is a qualitative methodology appropriate?	Yes / Can't tell / No
Is it worth continuing?	
Was the research design appropriate to address the aims of the research?	Yes / Can't tell / No
Was the recruitment strategy appropriate to address the aims of the research?	Yes / Can't tell / No
Was the data collected in a way that addressed the research question?	Yes / Can't tell / No
Has the relationship between researcher and participants been adequately considered?	Yes / Can't tell / No
Have ethical issues been taken into consideration?	Yes / Can't tell / No
Was the data analysis sufficiently rigorous?	Yes / Can't tell / No
Is there a clear statement of findings?	Yes / Can't tell / No
How valuable is the research?	

SOURCE: Critical Appraisal Skills Programme (2013). "Qualitative Checklist." Oxford, United Kingdom. Accessed from: http://media.wix.com/ugd/dded87_342758a916222fedf6e2355e17782256.pdf.

Study Name	<i>Dahal</i>	<i>Kabeer</i>	<i>Kilby</i>	<i>Knowles</i>	<i>Kumari</i>	<i>Sahu</i>	<i>Maclean</i>	<i>Mathrani</i>	<i>Mercer</i>	<i>Pattenden</i>	<i>Ramachandar</i>
Was there a clear statement of the aims of the study?	yes	yes	Yes	yes	Yes	yes	Yes	Yes	Yes	Yes	yes
Is the qualitative methodology appropriate?	yes	yes	Yes	yes	Yes	yes	Yes	Yes	Yes	Yes	yes
Was the research design appropriate to address the aims of the research?	yes	yes	Yes	yes	Yes	yes	Yes	Yes	Yes	Yes	yes
Was the recruitment strategy appropriate to the aims of the research?	yes	yes	Yes	yes	Yes	no	Yes	Yes	Yes	Yes	yes
Was the data collected in a way that addressed the research issue?	yes	yes	Yes	yes	Yes	yes	Yes	Yes	Yes	Yes	yes

Study Name	<i>Dahal</i>	<i>Kabeer</i>	<i>Kilby</i>	<i>Knowles</i>	<i>Kumari</i>	<i>Sahu</i>	<i>Maclean</i>	<i>Mathrani</i>	<i>Mercer</i>	<i>Pattenden</i>	<i>Ramachandar</i>
Has the relationship between researcher and participants been adequately considered?	yes	Can't Tell	Can't Tell	yes	yes	Can't Tell	yes	Can't tell	No	Can't tell	Can't Tell
Have ethical issues been taken into consideration?	yes	Can't Tell	No	yes	No	Can't Tell	No	Can't tell	No	No	yes
Was the data analysis sufficiently rigorous?	yes	yes	Yes	yes	Yes	yes	Yes	Yes	Yes	Yes	yes
Is there a clear statement of findings?	yes	yes	Yes	yes	Yes	yes	Yes	Yes	Yes	Yes	yes
Is the research valuable?	Valuable	Valuable	Valuable	Valuable	Valuable	Valuable	Valuable	Valuable	Valuable	Valuable	Valuable

Appendix 6: Quantitative risk of bias assessment tool

Code description	Code	Comment
Study ID	Last name of author, year	
Justification of use	Study design and methodology	
Ask these questions for all quantitative studies		
Does the study show baseline characteristics of beneficiaries and non-beneficiaries?	1 = Yes 2 = No 9 = Unclear 99 = Not applicable	Comment: Open answer
If baseline characteristics are not available, does the study show characteristics of beneficiaries and non-beneficiaries that are not likely to be affected by the intervention?		
Are the mean values or the distributions of the covariates at baseline statistically different for beneficiaries and non-beneficiaries (p<0.05)		
If there are statistically significant differences in plausibly exogenous characteristics between beneficiaries and non-beneficiaries are these differences controlled for using covariate analysis in the impact evaluation?		
If baseline characteristics are not available, does the study qualitatively assess why beneficiaries are likely/unlikely to be a random draw of the population at baseline?		
Confounding and selection bias (ask questions for all quantitative studies)		
Does the study use a comparison/control group of students/households without access to the program?	1 = Yes 2 = No 9 = Unclear 99 = Not applicable	Comment: Open answer
Does the study use a comparison/control group of students/households with access to the program but without participation in the program?		
Does the study include data at baseline and endline (before and after the intervention)?		
Are the data on covariates collected at the baseline?		
Is difference in differences estimation (i.e. using statistical inference) used?		
If the study is quasi-experimental and uses difference-in-difference estimation do the authors assess the parallel trends assumption?		
If the study does not use difference in difference, does the study control for baseline values of the outcome of interest		

If the study does not use difference in difference and does not control for baseline values of the outcome variable, does the study control for other covariates at baseline		
If the study does not use difference in differences estimation, is there any assessment of likely risk of bias from time invariant characteristics driving both participation and outcome?		
If the study does not use difference in difference estimation but does assess likely risk of bias from time invariant characteristics, are these time invariant characteristics likely to bias the impact estimates		
Does the study report the table with the results of the outcome equation (including covariates)? Where full results of the outcome equation are not reported, is it clear which covariates have been used?		
Are all relevant observable covariates (confounding variables) included in the outcome equation which might explain outcomes, if estimation does not use a statistical technique to control for selection bias (RCT, PSM or covariate matching, IV or switching regression)? This might, for example, include control for ability, and/or social capital.		
Attrition (ask questions for all quantitative studies)		
For studies including baseline data, does the study report attrition (drop-out) rates?	1 = Yes 2 = No	Comment: Open answer
Is the attrition rate below 10% ?	9 = Unclear	
Does the study assess whether drop-outs are random draws from the sample (e.g. by examining correlation with determinants of outcomes, in both treatment comparison group)?	99 = Not applicable	
Spillovers and contamination (ask questions for all quantitative studies)		
Spillovers: are comparisons sufficiently isolated from the intervention (e.g., participants and non-participants are sufficiently geographically or socially separated) or are spillovers estimated by comparing non-beneficiaries with access to the intervention to non-beneficiaries without access to the intervention and/or through social network analysis?	1 = Yes 2 = No 9 = Unclear 99 = Not applicable	Comment: Open answer
Spillovers; if spillovers are not estimated, is the study likely to bias the impact of the program?		
Contamination: does the study assess whether the control group receives the intervention?		
Contamination: if the control group receives the intervention but for a shorter amount of time does the study assess the likelihood		

that the control group has received equal benefits as the treatment group		
Contamination: if the control group receives the intervention have they received the intervention sufficiently long to argue that they have benefited from the intervention		
Contamination: does the study describe and control for other interventions which might explain changes in outcomes?		
Other threats to validity (ask questions for all quantitative studies)		
Does the evidence suggest analysis reporting biases are a serious concern? Analysis reporting biases include failure to report important treatment effects (possibly relating to intermediate outcomes), or justification for (uncommon) estimation methods, especially multivariate analysis for outcomes equations.	1 = Yes 2 = No 9 = Unclear 99 = Not applicable	Comment: Open answer
Are there concerns about baseline data collected retrospectively		
Are there concerns about courtesy bias, social acceptability bias, political correctness bias, self-serving bias, self-importance bias and biases in reporting of sensitive information from outcomes collected through self-reporting?		
Construct Validity (ask questions for all quantitative studies)		
Was the survey suitable for the local context?	1 = Yes 2 = No 9 = Unclear 99 = Not applicable	Comment: Open answer
Does the study describe the implementation of the program in sufficient detail?		
Does the study take into consideration potential implementation failures		
Does the study use a proper theory of change, logframe and/or other proper conceptual or theoretical framework?		
Does the study analyze the outcome measures put forward in the theory of change or logframe?		
Was the implementation of the intervention influenced by the research?		
Did the researchers have perfect control over the intervention?		
Was the implementing agency representative for the agencies that usually implement self-help group programs?		
External Validity (ask questions for all quantitative studies)		
Is the study sample representative of the population of interest?	1 = Yes	Comment: Open answer

Was the effectiveness of the intervention harmed by implementation failures that would not have happened in the absence of the research?	2 = No 9 = Unclear 99 = Not applicable	
Does the study assess the replicability of the intervention?		
Is the intervention replicable?		
Does the study assess the scalability of the intervention?		
Is the intervention scalable?		
Do the authors clearly distinguish between the intention-to-treat effect and the treatment effect on the treated?		
Do the authors highlight the intention-to-treat effect?		
Hawthorne and John Hendry Effects (ask questions for all quantitative studies)		
Do the authors argue convincingly that it is not likely that being monitored influences the behavior of the beneficiaries and non-beneficiaries in different ways?	1 = Yes 2 = No 9 = Unclear 99 = Not applicable	Comment: Open answer
Confidence Intervals (ask questions for all quantitative studies)		
Does the study account for lack of independence between observations within assignment clusters if the outcome variables are clustered?	1 = Yes 2 = No 9 = Unclear 99 = Not applicable	Comment: Open answer
Is the sample size likely to be sufficient to find significant effects of the intervention?		
Do the authors control for heteroskedasticity and/or use robust standard errors?		
Ask questions below only for studies that apply randomization		
Does the study apply randomized assignment?	1 = Yes 2 = No 9 = Unclear 99 = Not applicable	Comment: Open answer
Does the study use a unit of allocation with a sufficiently large sample size to ensure equivalence between the treatment and the control group?		
Ask questions below only for studies that apply regression discontinuity designs		
Is the allocation of the program based on a pre-determined continuity on a continuous variable and blinded to the beneficiaries or if not blinded, individuals cannot reasonably affect the assignment variable in response to knowledge of the participation rule?	1 = Yes 2 = No 9 = Unclear	Comment: Open answer

Is the sample size immediately at both sides of the cut-off point sufficiently large to equate groups on average?	99 = Not applicable	
Is the mean of the covariates of individuals immediately at both sides of the cut-off point statistically significantly different for beneficiaries and non-beneficiaries?		
If there are statistically significant differences between beneficiaries and non-beneficiaries are these differences controlled for using covariate analysis?		
Ask questions below only for studies that apply matching		
Quality of matching (PSM, covariate matching)		
Are beneficiaries and non-beneficiaries matched on all relevant characteristics?	1 = Yes 2 = No 9 = Unclear 99 = Not applicable	Comment: Open answer
Does the study report the results of the matching function (e.g. for PSM the logit function)?		
Does the study report the matching method?		
Does the study exclude observations outside the common support?		
Does the study use variables at follow-up that can be affected by the intervention in the matching equation?		
Are matches found for the majority of participants (>90%)?		
If >=10% of participants failed to be matched, is sensitivity analysis used to re-estimate results using different matching methods?		
For nearest-neighbor PSM, does the study report the mean or distribution of the propensity scores in the treatment and control groups after matching?		
For nearest-neighbor PSM, are propensity scores similar, based on tests for statistical differences at the means or other quantiles of the distribution)?		
Does the study report the mean or distribution for the covariates of the treatment and control groups after matching?		
Are these characteristics similar, based on tests for statistically significant differences ($p>0.5$)?		
Do the authors use bootstrapped standard errors?		
Sensitivity analysis (only for studies that apply PSM)		
For PSM, where propensity score distributions and/or covariates of the treatment and control groups are not reported, or they are reported but there are differences in means or distributions of the covariates or propensity scores (usually only applicable to methods which do not exclude treatment	1 = Yes 2 = No 9 = Unclear	Comment: Open answer

observations such as nearest neighbor), is robustness assessed using an additional matching technique?	99 = Not applicable	
Is sensitivity to hidden bias assessed statistically, e.g., using the Rosenbaum bounds test?		
Ask questions below only for studies that apply instrumental variable estimation		
Quality of IV, two-steps endogenous switching regression approach		
Does the study describe clearly the instrumental variable(s)/identifier used?	1 = Yes 2 = No 9 = Unclear 99 = Not applicable	Comment: Open answer
Are the results of the participation equation reported?		
Are the instruments jointly significant at the level of $F \geq 10$? If an F test is not reported, does the author report and assess whether the R-squared of the instrumenting equation is large enough for appropriate identification ($R\text{-sq} > 0.5$)		
Are the instruments individually significant ($p \leq 0.05$)?		
For IV, If more than one instrument is used in the procedure, does the study include and report an overidentifying test ($p \leq 0.05$ is required to reject the null hypothesis)?		
Does the study qualitatively assess the exogeneity of the instrument/identifier (both externality as well as why the variable should not enter by itself in the outcome equation)?		
Ask questions below only for studies with censored outcome variables		
Do the authors use appropriate methods (e.g. Heckman selection models, tobit models, duration models) to account for the censoring of the data?	1 = Yes 2 = No 9 = Unclear 99 = Not applicable	Comment: Open answer
For Heckman models; is there is a variable that is statistically significant in the first stage of the selection equation and excluded from the second stage		
Overall Assessment		
Assessment Selection Bias	Low risk of bias	Comment: Open answer
Assessment Spillovers and Contamination Bias	Medium risk of bias	
Assessment Outcome and Analysis Reporting Bias	High risk of bias	
Assessment Other biases	Unclear risk of bias	

Appendix 7: Overview of risk of bias assessment of included quantitative studies

	Selection Bias and Confounding	Performance Bias: Assessment Spillovers and Contamination	Outcome and Analysis Reporting Biases	Other Biases
Ahmed, 2005	High risk of bias	High risk of bias	Medium risk of bias	Medium risk of bias
Swain and Wallentin, 2009	High risk of bias	High risk of bias	High risk of bias	High risk of bias
Banerjee et al., 2015, 2010	Low risk of bias	Medium risk of bias	Low risk of bias	Low risk of bias
Coleman, 1999	High risk of bias	High risk of bias	Low risk of bias	Low risk of bias
De Hoop et al., 2014	Medium risk of bias	High risk of bias	Low risk of bias	Medium risk of bias
Deininger and Liu, 2013, 2009	Medium risk of bias	Low risk of bias	Medium risk of bias	High risk of bias
Desai and Joshi, 2012	Low risk of bias	Low risk of bias	Low risk of bias	Low risk of bias
Desai and Tarozzi, 2011	Low risk of bias	Low risk of bias	Low risk of bias	Medium risk of bias
Garikipati 2012, 2008	High risk of bias	High risk of bias	High risk of bias	Low risk of bias
Holvoet, 2005	High risk of bias	Medium risk of bias	Low risk of bias	Medium risk of bias
Husain et al., 2010	High risk of bias	High risk of bias	High risk of bias	High risk of bias
Kim et al., 2009 and Pronyk et al., 2006	Medium risk of bias	Low risk of bias	Medium risk of bias	Medium risk of bias
Kundu, et al., 2011	High risk of bias	High risk of bias	High risk of bias	High risk of bias
Mahmud, 1994	High risk of bias	High risk of bias	High risk of bias	Medium risk of bias

Nessa et al., 2012	High risk of bias	High risk of bias	High Risk of Bias	Low risk of bias
Osmani, 2007	High risk of bias	High risk of bias	High risk of bias	Low risk of bias
Pitt et al., 2006	Medium risk of bias	High risk of bias	Medium risk of bias	Low risk of bias
Sherman et al., 2010	Medium risk of bias	High risk of bias	High risk of bias	High risk of bias
Steele et al., 1998	High risk of bias	High risk of bias	High risk of bias	Low risk of bias
Swendeman et al., 2009	High risk of bias	Low risk of bias	Medium risk of bias	Medium risk of bias

Appendix 8: Detailed risk of bias assessment of included quantitative studies

	Selection Bias and Confounding	Performance Bias: Assessment, Spillovers, and Contamination	Outcome and Analysis Reporting Biases	Other Biases
Ahmed, 2005	High risk of bias The study does not adequately control for selection bias in the analysis.	High risk of bias The study does not take into consideration that the comparison group may also have been contaminated by the intervention.	Medium risk of bias The study assesses the impact of several components of the intervention without taking into consideration selection bias. This is an uncommon estimation method, which suggests that the analysis is vulnerable to analysis reporting biases. The study also uses co-variables in the model that may be endogenous.	Medium risk of bias The answers to the questions about domestic violence are vulnerable to social desirability bias.
Bali Swain and Wallentin, 2009	High risk of bias The study uses analysis to separately determine the trend of the outcome measures among the beneficiaries and the non-beneficiaries. This does not allow for estimating the impact of the intervention. Hence, the study does not use a valid identification strategy	High risk of bias The study selects the comparison group from the same location as the beneficiaries so there is a potential for spillovers biasing the findings.	High risk of bias The study uses an unusual type of analysis (separately determining the trend for the beneficiaries and the non-beneficiaries). This could bias the research findings.	High risk of bias The use of recall data could bias the impact estimates. And it is not well explained why the use of these data can be considered valid for this study.
Banerjee et al., 2015, 2010	Low risk of bias The study uses a matched-pair cluster-randomized controlled trial. Baseline and follow-up data were collected, but panel data were not available (i.e., the respondents in the follow-up are not necessarily the same as the respondents at baseline due to resampling). The study assesses equivalence of treatment and control areas at baseline and endline and does not find significant differences. The study controls for clustering in the calculation of the standard errors. Other microfinance interventions were rolled out during the study period in both	Medium risk of bias The control group was contaminated by other types of microfinance. The study notes that other microfinance interventions were rolled out during the study period in both treatment and control areas and notes that both treatment and comparison areas were borrowing microcredit (thought	Low risk of bias There do not appear to be serious outcome or analysis reporting biases.	Low risk of bias There do not appear to be serious other biases. The outcome measures do not seem to be vulnerable to social desirability bias.

	treatment and control areas. Both treatment and comparison areas were borrowing microcredit (though borrowing rates were lower in the comparison area). However, the study does not control for the other microfinance interventions in the analysis. Instead, the authors calculate an intention-to-treat effect.	borrowing rates were lower in comparison area). However, the study does not control for the other intervention effects in the analysis.		
Coleman, 2002	High risk of bias	High risk of bias	Low risk of bias	Low risk of bias
	The study uses a multivariate regression model that includes a dummy for participation in a self-help group, and a variable capturing the number of months during which self-help group members received credit as explanatory variables. Although the authors claim that this methodology allows for controlling for selection bias, this methodology cannot be considered a credible identification strategy. First, it is not clear how the beneficiaries of the intervention were selected. Second, there may have been self-selection among those beneficiaries who started benefiting from the intervention at an early stage. Third, the study does not control for selection bias based on unobservables. These problems cannot be resolved by including village fixed effects.	The comparison group includes non-beneficiaries who could have been affected by the intervention due to their close proximity to the beneficiaries of the intervention. Hence, the findings of the evaluation could be biased due to spillovers.	There do not appear to be serious outcome or analysis reporting biases.	There do not appear to be serious other biases. The outcome measures do not seem to be vulnerable to social desirability bias.
De Hoop et al., 2014	Medium risk of bias	High risk of bias	Low risk of bias	Medium risk of bias
	The study uses a propensity score matching design without baseline data. For the nearest neighbor matching, the study does not report the mean or distribution of the propensity scores in the treatment and control groups after matching, or the mean or distribution for the covariates of the treatment and control groups after matching, but the study does control for robustness of the results using kernel matching.	There is a potential bias from spillover effects as the non-members (the comparison group) are drawn from the same villages as the SHG members (treatment group).	There do not appear to be serious outcome or analysis reporting biases now that the authors have responded with a set of analyses with additional outcome measures.	The answers to the questions about domestic violence are vulnerable to social desirability bias.

Deininger and Liu, 2013, 2009	Medium risk of bias	Low risk of bias	Medium risk of bias	High risk of bias
	The study uses propensity score matching and difference-in-difference estimation. The study uses recall data over a four-year recall period for the DID estimation component. This may result in bias.	The authors estimate a combined intention-to-treat effect for women who decide to self-select into SHGs and women who decide not to self-select into SHGs. This minimizes the risk of spillovers.	The two versions of this paper report slightly different results, which may indicate outcome reporting bias. Standard deviations are not reported in either version of the paper, and authors did not respond to requests for information. As a result, the standard deviations had to be imputed increasing the potential risk of bias of the effect size.	The use of recall data could result in social desirability bias in the measurement of empowerment.
Desai and Joshi, 2012	Low risk of bias	Low risk of bias	Low risk of bias	Low risk of bias
	It appears that the randomization resulted in balance across observable and unobservable characteristics.	The authors estimate a combined intention-to-treat effect for women who decide to self-select into SHGs and women who decide not to self-select into SHGs. This minimizes the risk of spillovers.	There do not appear to be serious outcome or analysis reporting biases.	There do not appear to be other serious biases.
Desai and Tarozzi, 2011	Low risk of bias	Low risk of bias	Low risk of bias	Medium risk of bias
	The randomized controlled trial design suggests there is balance across observable and unobservable characteristics, and the balance test suggests that the randomization has worked, although not perfectly due to noncompliance. Nonetheless, the researchers choose a valid instrumental variable approach to account for the noncompliance with the randomization. The study collects baseline and endline data at the individual level, but not for the same individuals. The authors therefore estimate mean effects at the village level, thus considerably reducing the power of the study—the sample size is only 54 PAs in Amhara, and 78 PAs in Oromia, which may be insufficient to detect small/medium-sized effects.	There do not appear to be serious concerns about spillovers.	There do not appear to be serious concerns about outcome or analysis reporting biases.	Condom use is a sensitive variable. This could increase measurement error. This is not discussed in the paper.

Garikipati, 2012, 2008	However, the authors control for several plausibly exogenous control variables, which should normally increase the statistical power of the study.			
	High risk of bias	High risk of bias	High risk of bias	Low risk of bias
	The study uses a cross-sectional design and instrumental variables estimation to address selection bias. The validity of the instrumental variable is not discussed or tested, which increases the risk of bias. The study does not measure covariates or outcomes at baseline. It does not take into account clustering in the analysis and does not report the use of cluster-robust standard errors. The authors include the "own use of loan" as an explanatory variable. This intermediate outcome variable should not have been included in the outcome equation.	Spillovers can bias the findings of this study because the non-beneficiaries come from the same village and may also have been affected by the intervention.	The study from 2008 uses unusual methods to construct the outcome variables. This may result in outcome reporting biases. Furthermore, the use of intermediate outcome variables as explanatory variables could result in analysis reporting biases.	There do not appear to be serious other biases.
Holvoet, 2005	High risk of bias	Medium risk of bias	Low risk of bias	Medium risk of bias
	This is an ex post multivariate multinomial logistic regression study without a valid identification strategy. The study does not collect baseline data and elicits baseline characteristics using recall over long periods. The study attempts to "match" the programs that deliver the treatments under study, but does not match participants nor otherwise control for selection bias in the analysis. Furthermore, the study does not use a dummy variable for membership as the treatment variable but the time women are members of self-help groups. This type of analysis does not take into consideration the possibility of nonlinearities.	There was potential for spillover effects, but the study reports that the authors attempted to minimize these by not sampling non-beneficiaries with close connections to the beneficiaries.	It does not appear that there are serious outcome or analysis reporting biases.	The study relies on retrospective baseline data collection to a considerable extent.
Husain et al., 2010	High risk of bias	High risk of bias	High risk of bias	High risk of bias
	The study uses multivariate regression analysis without a valid identification strategy. The study compares new to old SHG members and does not collect	The new members and old members appear to be selected from the same locations,	The study does not report the numeric value of the correlation coefficients, only whether these are positive or	The municipalities from which the sample was selected were

	outcome or covariate data at baseline. This makes it impossible to reliably evaluate the effectiveness of the program and extract reliable effect sizes. The study does also not control for selection bias and does not take clustering into consideration in the calculation of the standard errors.	suggesting that bias resulting from spillovers is an important concern.	negative and statistically significant or not.	chosen by the implementing agency based on their successful performance, suggesting that the results may not be representative of the target population.
Kim et al., 2009, and	Medium risk of bias	Low risk of bias	Medium risk of bias	Medium risk of bias
Pronyk et al., 2006	It is not clear whether the randomization was successful. The randomization was based on a relatively small sample of four treatment and four control villages. This increases the likelihood of observable and unobservable differences between the treatment and the control group. Furthermore, it is unclear how comparable the villages with only microfinance are.	The risk of spillovers is minimized because the control villages do not have access to the intervention.	The two studies report slightly different results and sample sizes, which may indicate outcome reporting bias.	The authors note several potential limitations, including the possibility of Hawthorne or other reporting biases.
Kundu et al., 2011	High risk of bias	High risk of bias	High risk of bias	High risk of bias
	In the most recent version of this paper, the study uses a multinomial logit regression analysis without a valid identification strategy. The authors do not use baseline data and do not report the results of the analysis. In the earlier version of the paper, the study uses panel data and difference-in-differences analysis. However, the intervention already started before the baseline survey. This invalidates the parallel trends assumption. The authors do also not take clustering into consideration in the estimation of the standard errors and do not assess the potential biases in outcome measurement.	The study uses both nonparticipants from treatment villages and nonparticipants from control villages in their analysis without separately analyzing these, which could result in a bias due to spillovers. Furthermore, several members from the comparison group were members of self-help groups during the baseline survey, suggesting that they may also have been affected by the intervention.	In the most recent version of this paper, the study does not report the results of the multinomial logit model. In the earlier version of the paper, the study is vulnerable to analysis reporting biases because of the start of the intervention before the collection of baseline data.	The baseline data were collected retrospectively, asking participants to recall information from two years ago.

Mahmud, 1994	High risk of bias	High risk of bias	High risk of bias	Medium risk of bias
	This is a cross-sectional study using multi-variate analysis without baseline data collection. The study does not use a valid identification strategy. The authors also do not take into consideration clustering in the calculation of the standard errors.	The control group is drawn from the same locations as the beneficiaries. Hence, the estimate of the impact of the intervention may be biased due to spillovers.	The study only assesses the impact of the program on two primary outcomes defined by the study. The authors decided not to analyze the effect of the program on the use of temporary contraception because "the bivariate frequency distributions have revealed that both the level and pattern of use of temporary methods was largely undifferentiated between the two [treatment and control] groups." However, there were significant differences in the characteristics of the treatment and control group, so one cannot apriori assume that an absence of a difference in the unadjusted outcome necessarily translates into an absence of effect following adjustment for confounding factors. And there may be a potential for outcome reporting bias.	There do not appear to be serious concerns about other biases. It is unclear whether the outcome variable measures empowerment.
Nessa et al., 2012	High risk of bias	High risk of bias	High Risk of bias	Low risk of bias
	The study uses a cross-sectional study design without data collection at baseline and does not have a valid identification strategy. The study only controls for a small number of potential confounding variables but also includes annual income, a variable likely affected by the program, as a control variable in the regression analysis.	The non-beneficiaries come from the same locality as the beneficiaries, which could invalidate the results due to spillovers.	The outcome variables are not well explained. It controls for a small number of potential confounding variables but also includes annual income, a variable likely affected by the program, as a control variable in the regression analysis.	There do not appear to be serious concerns about other biases.
	High risk of bias	High risk of bias	High risk of bias	Low risk of bias

Osmani, 2007	The study uses an instrumental variable approach to address the problem of selection bias. However, the validity of the instruments depends on the inclusion of household income as an independent variable, which is an intermediate outcome. The study also instruments for household income, but includes the participation variable in the equation that estimates household income. Thus, the study uses one endogenous variable to predict the other endogenous variable and vice versa, which suggests the instruments are not valid. Household income, being an intermediate outcome, should not be included in the model. The sample size is also too small to determine precise effects (42 treatment and 42 comparison households), and the study does not adjust for clustering.	The beneficiaries and the comparison group were drawn from the same villages. Hence, the estimates may be biased due to spillovers.	It appears that the authors do not apply the use of instrumental variables in a correct manner. This suggests that the findings are vulnerable to analysis reporting biases.	There do not appear to be serious other biases.
Pitt et al., 2006	Medium risk of bias The study uses a fixed-effects instrumental variable regression approach (and compares the results to ordinary least squares with village-level variables and fixed-effects estimation). The authors identify a set of instrumental variables and control for village-level fixed unobserved characteristics. It is unclear whether the instruments are valid.	High risk of bias The authors do not discuss the potential bias from spillover effects, even though the comparison women come from the same communities.	Medium risk of bias The study does not report the participating equation; it is unclear whether the instruments were jointly or independently significant; and the authors do not report a test for overidentification.	Low risk of bias There do not appear to be serious outcome and analysis reporting biases
	Medium risk of bias	High risk of bias	High risk of bias	High risk of bias

Sherman et al., 2010	The study is a randomized controlled trial. The authors adjust for age and household income at baseline in the multivariate analysis, but only for the analysis concerned with the self-reported number of sex exchange partners, which is the primary outcome of interest. The sample size (50 treatment and 50 control group members) is arguably insufficient to ensure equivalence between samples through randomization and underpowered to detect small to medium effects.	Correspondence with the authors suggests that the control group comes from the same community, which increases the vulnerability of the study to bias from spillovers. Furthermore, the participants in the treatment group received a cash transfer so it is not very clear whether the effect is really due to self-help groups.	It does not appear that there are serious outcome or analysis reporting biases. The study controls for different control variables for different outcomes, suggesting potential analysis reporting bias.	The study uses recall data for sensitive outcome measures, which could invalidate the results of the impact evaluation.
Steele et al., 1998	High risk of bias The study uses multivariate regression and controls for baseline characteristics as well as baseline values of the outcomes of interest. However, the study also controls for several potentially endogenous variables, which could result in a bias in the impact estimates. The study does not use a valid identification strategy.	High risk of bias The study compares beneficiaries to eligible non-beneficiaries in the same communities, so the results could be biased due to spillovers	High risk of bias There are serious inconsistencies in the reporting (the results reported in the text do not match those reported in the tables) and the authors only report results for some of the analyzed comparisons. The authors also mention analyses that are not reported in the study.	Low risk of bias There do not appear to be serious other biases. One variable was collected using recall (worked for cash or kind during last year), but this is unlikely to be a serious concern.
Swendeman et al., 2009	High risk of bias The study uses multivariate regression analysis, but does not use a valid identification strategy. The study compares randomly selected participants in a town that received the intervention to randomly selected participants in a town that received the control intervention, but does not discuss whether the intervention and treatment town were comparable, does not establish equivalence of treatment and comparison group participants, and does not appropriately control for selection bias.	Low risk of bias The comparison group seems sufficiently far away to mitigate concerns over bias from spillovers	Medium risk of bias There do not appear to be serious outcome and analysis reporting biases. Some outcome variables were not discussed because the authors do not find significant effects.	Medium risk of bias It looks like the research team influenced the fidelity of the intervention.

Appendix 9: Quality assessment for included qualitative studies

Study Name	Dahal 2014	Kabeer 2011	Kilby 2011	Knowles 2014	Kumari 2011	Maclean 2012	Mathrani 2006	Mercer 2002	Pattenden 2011	Ramachandar 2009	Sahu 2012
Was there a clear statement of the aims of the study?	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Is the qualitative methodology appropriate?	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Was the research design appropriate to address the aims of the research?	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Was the recruitment strategy appropriate to the aims of the research?	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Was the data collected in a	can't tell	can't tell	yes	yes	yes	yes	yes	yes	yes	yes	yes

way that addressed the research issue?											
Has the relationship between researcher and participants been adequately considered?	no	can't tell	yes	yes	can't tell	can't tell	can't tell	can't tell	can't tell	can't tell	can't tell
Have ethical issues been taken into consideration?	yes	can't tell	can't tell	yes	can't tell	can't tell	can't tell	can't tell	can't tell	yes	can't tell
Was the data analysis sufficiently rigorous?	yes	yes	yes	yes	can't tell	can't tell	yes	can't tell	yes	yes	can't tell
Is there a clear statement of findings?	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Based on the above, is the research valuable?	valuable	valuable	valuable	valuable	valuable	valuable	valuable	valuable	valuable	valuable	valuable

Appendix 10: Procedures for calculating effect sizes

This appendix describes the procedure for calculating the effect sizes of the included quantitative studies.

First, we calculated standardized mean differences (Cohen's d) by dividing the mean difference with the pooled standard deviation by applying the formula in Equation 10.1:

$$(10.1) \text{ SMD} = \frac{Y_t - Y_c}{S_p}$$

Here, SMD refers to the standardized mean differences, Y_t refers to the outcome for the treatment group, Y_c refers to the outcome for the control or comparison group, and S_p refers to the pooled standard deviation.

The pooled standard deviation S_p can be calculated or approximated (in regression studies) using the following two formulas in Equations 10.2 and 10.3:

$$(10.2) S_p = \sqrt{\frac{((SD_y)^2 * (nt + nc - 2)) - (\frac{\beta^2 * (nt * nc)}{nt + nc})}{nt + nc}}$$

$$(10.3) S_p = \sqrt{\frac{(nt - 1) * s_t^2 + (nc - 1) * s_c^2}{nt + nc - 2}}$$

Equation 10.2 was used for regression studies with a continuous dependent variable for which we had information about the point estimate for the treatment variable and the associated standard deviation. SD_y refers to the standard deviation for the point estimate from the regression, n_t refers to the sample size for the treatment group, n_c refers to the sample size for the control group, and β refers to the point estimate. Equation 10.3 was applied when there was information about the standard deviation for the treatment group and the standard deviation for the control group. In this formula, s_t refers to the standard deviation for the treatment group and s_c to the standard deviation for the control group. We assumed the same standard deviation for the treatment and the control or comparison group when the paper only reported the standard deviation for the full sample, treatment group, or control or comparison group.

Then we corrected the standardized mean difference for potential bias from a small sample size using the formula to transform Cohen's d to Hedges' g in Equation 10.4:

$$(10.4) \text{ SMD}_{\text{corrected}} = \text{SMD}_{\text{uncorrected}} * \left(1 - \frac{3}{4 * (nt + nc - 2) - 1}\right)$$

Finally, we calculated the standard error of the standardized mean difference using Equation 10.5:

$$(10.5) \text{ SE} = \sqrt{\frac{nt + nc}{nc * nt} + \frac{\text{SMD}^2}{2 * (nc + nt)}}$$

For dichotomous variables, we used odds ratios and log odds ratios rather than risk ratios because methods are available to convert the natural logarithm of odds ratios to the standardized mean difference and vice versa, as illustrated in the formula in Equation 10.6 (Borenstein et al., 2009):

$$(10.6) g = \text{LogOddsRatio} * \frac{\sqrt{3}}{\pi}$$

This transformation required several statistical assumptions but it allowed for one meta-analysis with both dichotomous and continuous variables for the same construct. Conducting one meta-analysis for dichotomous and continuous variables was preferable because it substantially increased the number of studies we could include in one meta-analysis.

It was also appropriate because the included studies that analyzed continuous variables shared goals in common with the included studies that analyzed dichotomous variables. Borenstein et al. (2009) suggests that the transformation of log odds ratios to standardized mean differences improves the meta-analysis as long as the outcome variables measure the same construct. It is less important whether the outcome variables use different measurement scales. Nonetheless, the transformation from log odds ratios to standardized mean differences requires several statistical assumptions (Borenstein et al., 2009).

Following the correction of the effect size, we estimated the corrected standard error by applying the formula in Equation 10.7 for standardized mean differences that were estimated from odds ratios:

$$(10.7) SE_{corrected} = \sqrt{\frac{nt+nc}{nt*nc} + \frac{(SMD_{corrected})^2}{2*(nt+nc)}}$$

To derive odds ratio from studies that applied linear probability models, we assumed linearity in the estimation of standardized effect sizes from the linear probability model. In practice, this meant that if we observed a mean baseline value for the comparison group of 0.067 and an effect size of 3.1 percentage points, then we assumed that the follow-up value for the treatment group would be 0.067+0.031=0.098 and we assumed that the follow-up value for the comparison group would be 0.067. Using this information, we were able to estimate odds ratios using a 2 by 2 contingency table (Lipsey & Wilson, 2001), as described in Figure 10.1:

Figure 10. 1: Estimation of odds-ratios

		Frequencies	
		Success	Failure
Beneficiaries	A		B
Comparison Group	B		D

From the figure, we calculated the odds-ratio using Equation 10.8 where $\overline{ES}$ refers to the effect size:

$$(10.8) \overline{ES} = \frac{ad}{bc}$$

We then calculated the standard error of the natural logarithm of the odds ratio by calculating the number of cases where the treatment group could be considered empowered and the number of cases where the control or comparison group could be considered empowered. We did this by using the information about the percentage of empowered women in the treatment and control or comparison group, and information about the sample size in the treatment and

control or comparison group. This allowed us to estimate the standard error of the natural logarithm of the odds ratio using the following formula in Equation 10.9, where n_{11} is the number of empowered women in the treatment group, n_{10} is the number of empowered women in the control group, n_{01} is the number of nonempowered women in the treatment group, and n_{00} is the number of nonempowered women in the control group.

$$(10.9) \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{10}} + \frac{1}{n_{01}} + \frac{1}{n_{00}}}$$

Then we converted the log-odds ratios and their 95 per cent confidence intervals back to odds ratios as well as to standardized mean differences using the formula to transform log odds ratios to standardized mean differences. Following this conversion, we converted the standardized mean difference (Cohen's d) to Hedges' g to account for potential bias from small samples using the formula in Equation 10.10 to correct for potential bias from a small sample size:

$$(10.10) \text{SMD}_{\text{corrected}} = \text{SMD}_{\text{uncorrected}} * \left(1 - \frac{3}{4*(nt+nc-2)-1}\right)$$

We were also able to estimate the variance and standard deviation of outcome variables for which the standard deviation was not reported but for which the full distribution was reported. For this purpose, we used the formula from Equation 10.11:

$$(10.11) SD(X) = \sqrt{\frac{\sum(x-\mu)^2}{n-1}}$$

Here, μ is the mean value of x and n is the number of observations.

In the absence of standard errors for the regression analysis, we estimated the standard error of the mean effect size by dividing the point estimate by the t -value that is associated with significance at the 90, 95, and 99 per cent significance level, respectively. This procedure ensured the estimation of conservative pooled standard deviations.

Appendix 11: Additional forest plots

Figure 11. 1: Meta-Analysis for determining the effect of women's self-help groups on women's economic empowerment based on quasi-experimental evaluations with a high risk of selection-bias

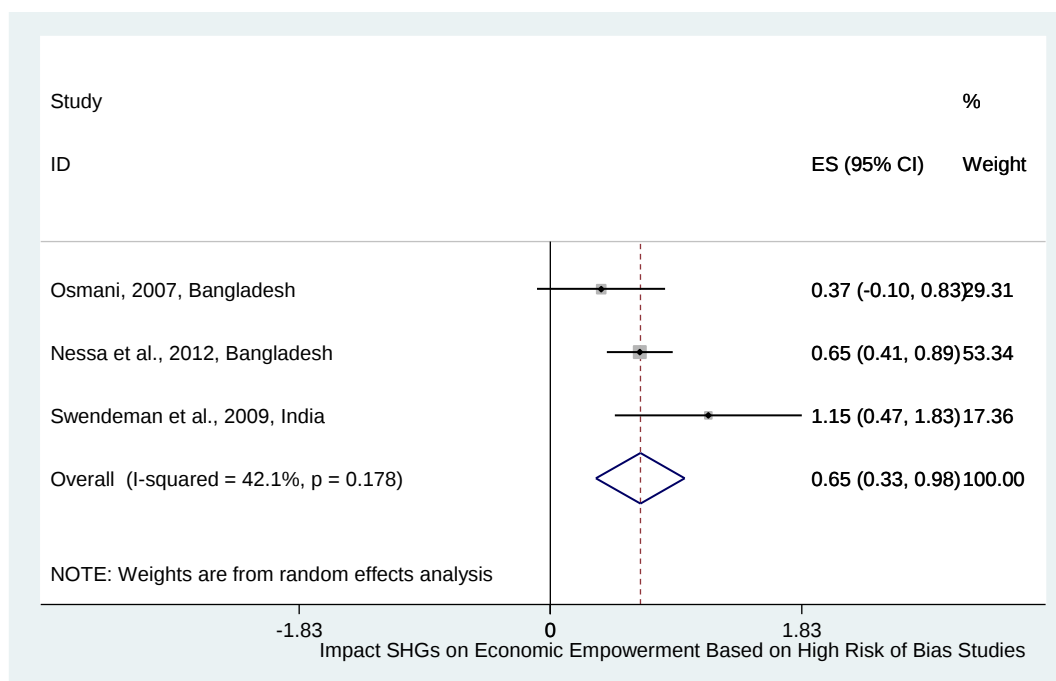


Figure 11. 2: Meta-Analysis for determining the effect of women's self-help groups on women's economic empowerment based on quasi-experimental evaluations with a medium risk of selection-bias

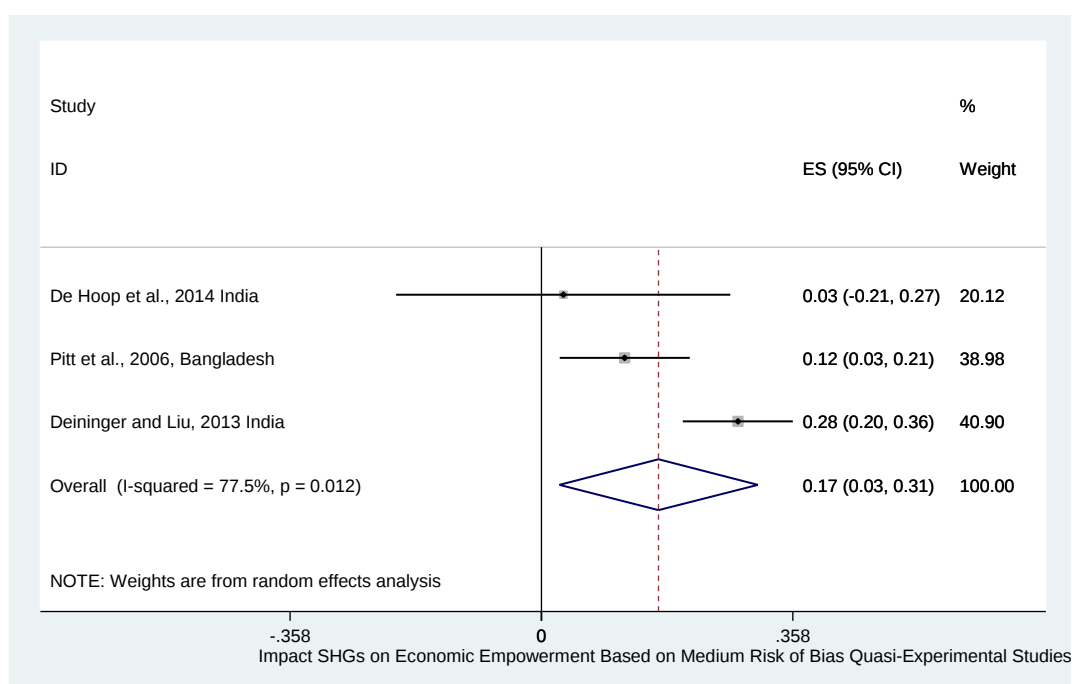


Figure 11. 3: Meta-Analysis for determining the effect of women's self-help groups on women's economic empowerment based on RCTs and quasi-experimental evaluations with a medium risk of selection-bias that focus on SHGs with a training component

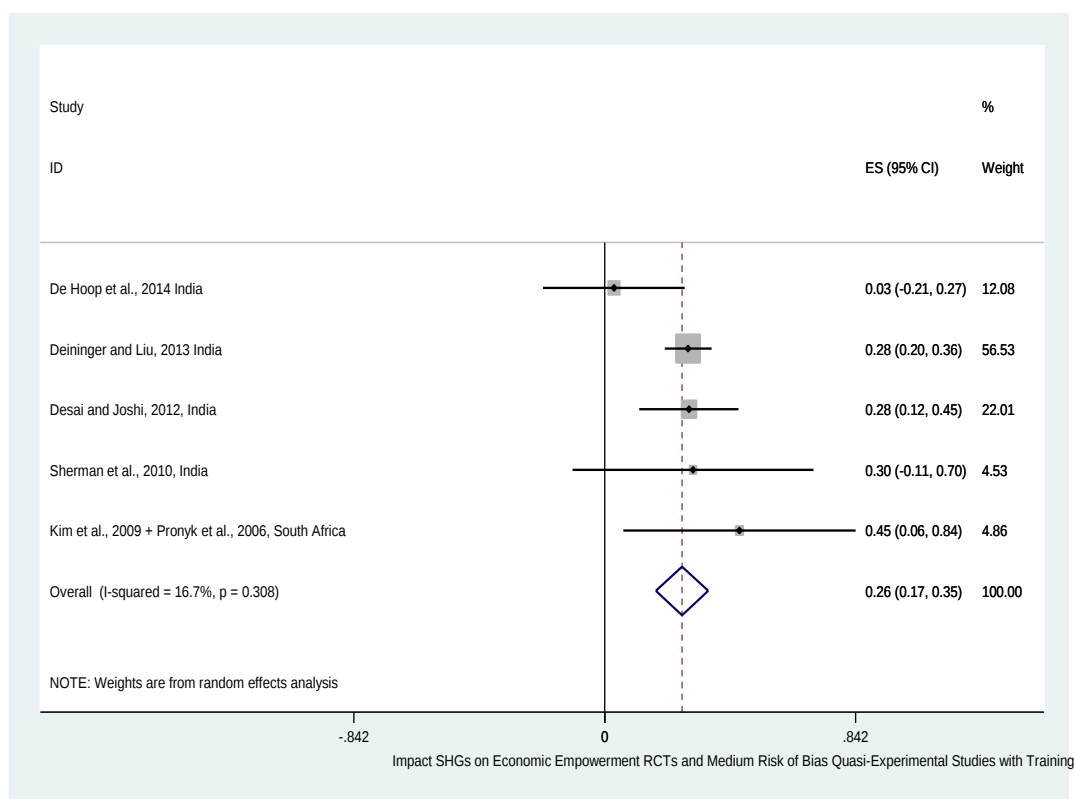


Figure 11. 4: Meta-Analysis for determining the effect of women's self-help groups on women's economic empowerment based on RCTs and quasi-experimental evaluations with a medium risk of selection-bias that focus on SHGs without a training component

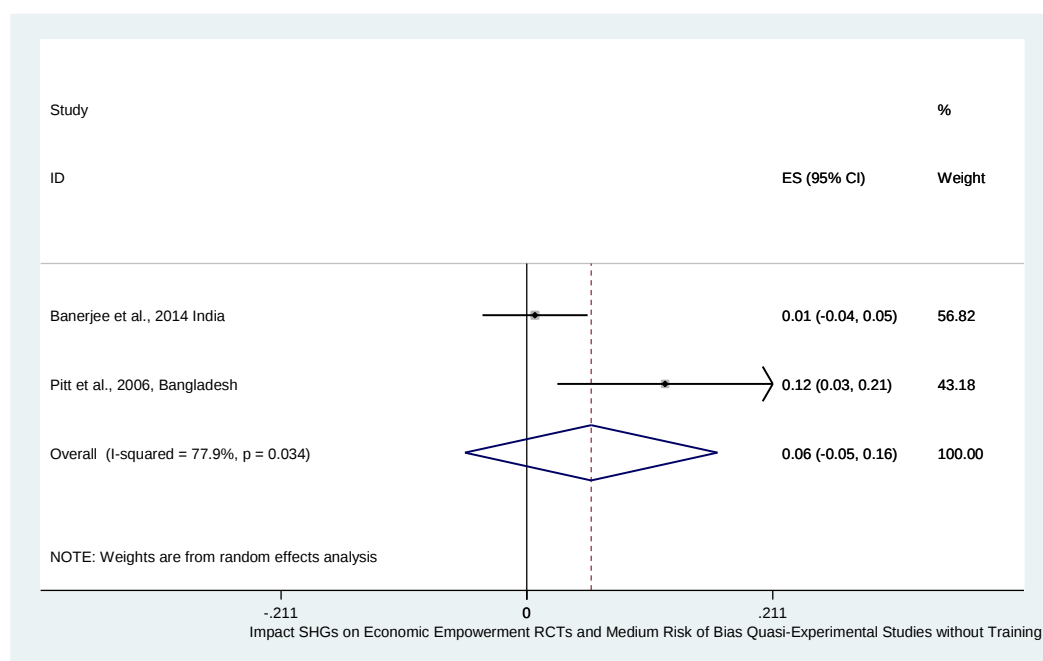


Figure 11. 5: Meta-Analysis for determining the effect of women's self-help groups on women's social empowerment based on quasi-experimental evaluations with a high risk of selection-bias

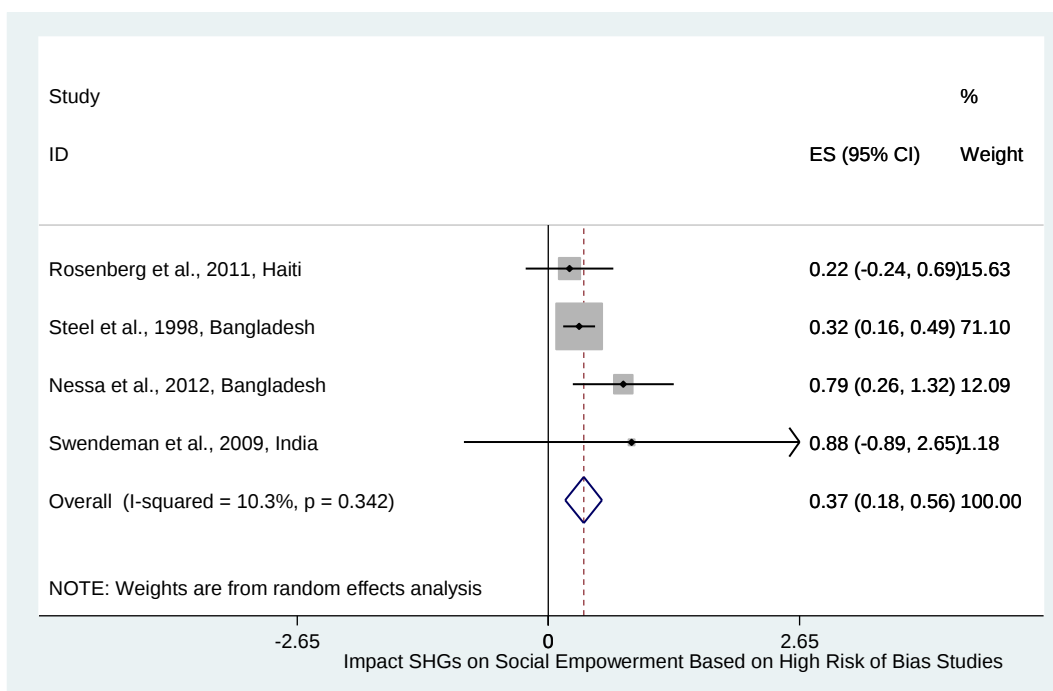


Figure 11. 6: Meta-Analysis for determining the effect of women's self-help groups on women's social empowerment based on quasi-experimental evaluations with a medium risk of selection-bias

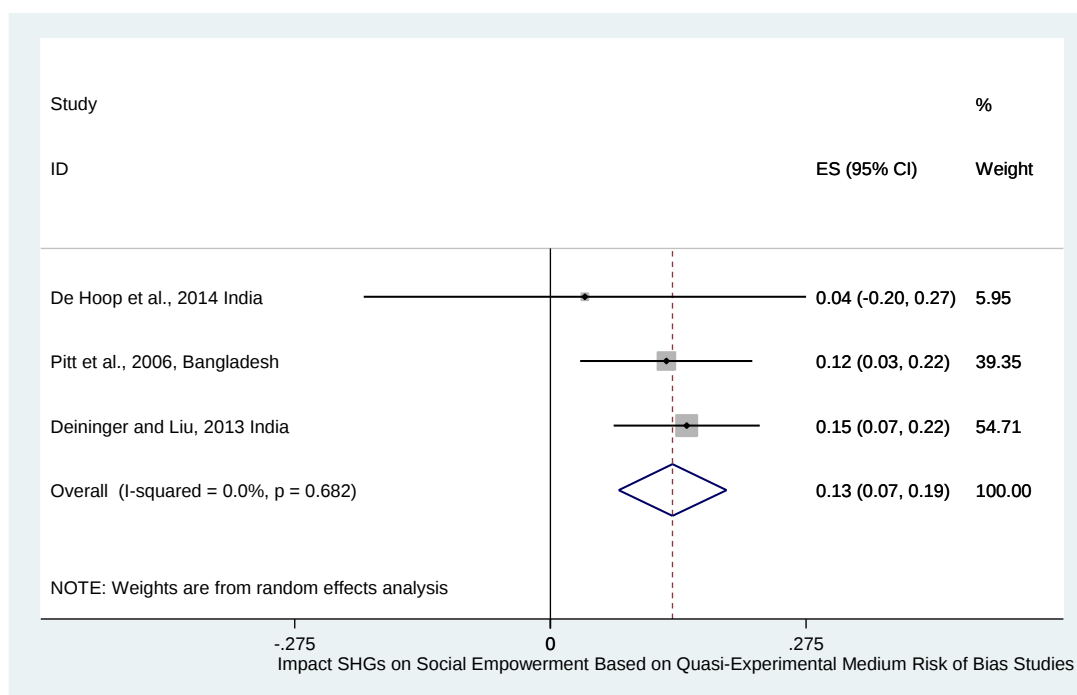


Figure 11. 7: Meta-Analysis for determining the effect of women's self-help groups on women's family-size decision making power based on quasi-experimental evaluations with a high risk of selection-bias

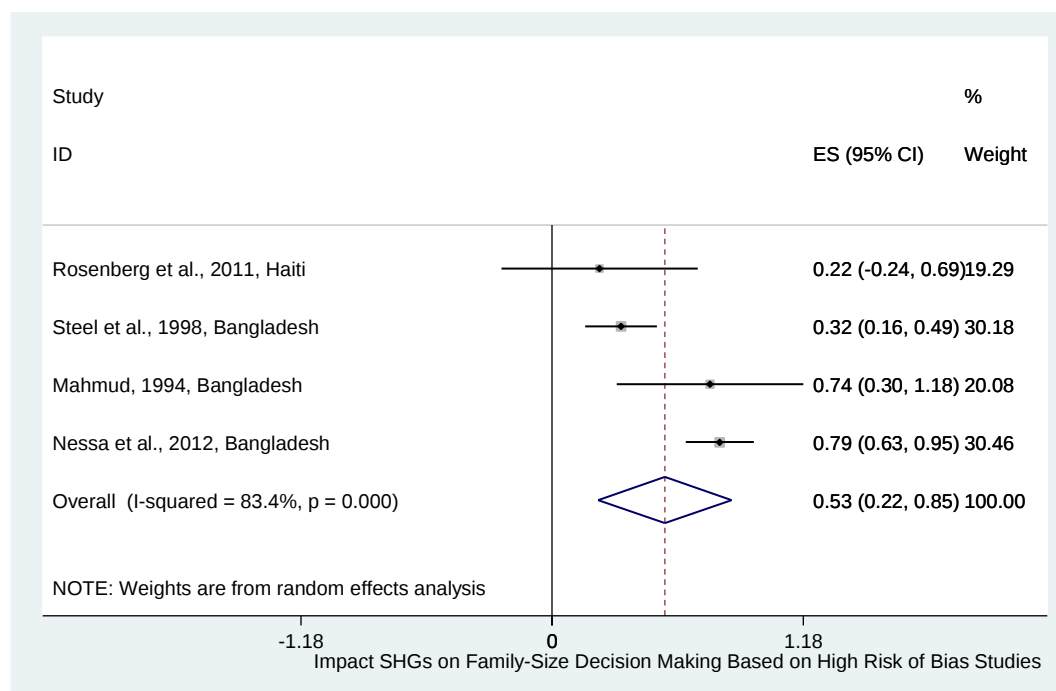


Figure 11. 8: Meta-Analysis for determining the effect of women's self-help groups on women's family-size decision-making based on RCTs and quasi-experimental evaluations with a medium risk of selection-bias that focus on SHGs with a training component

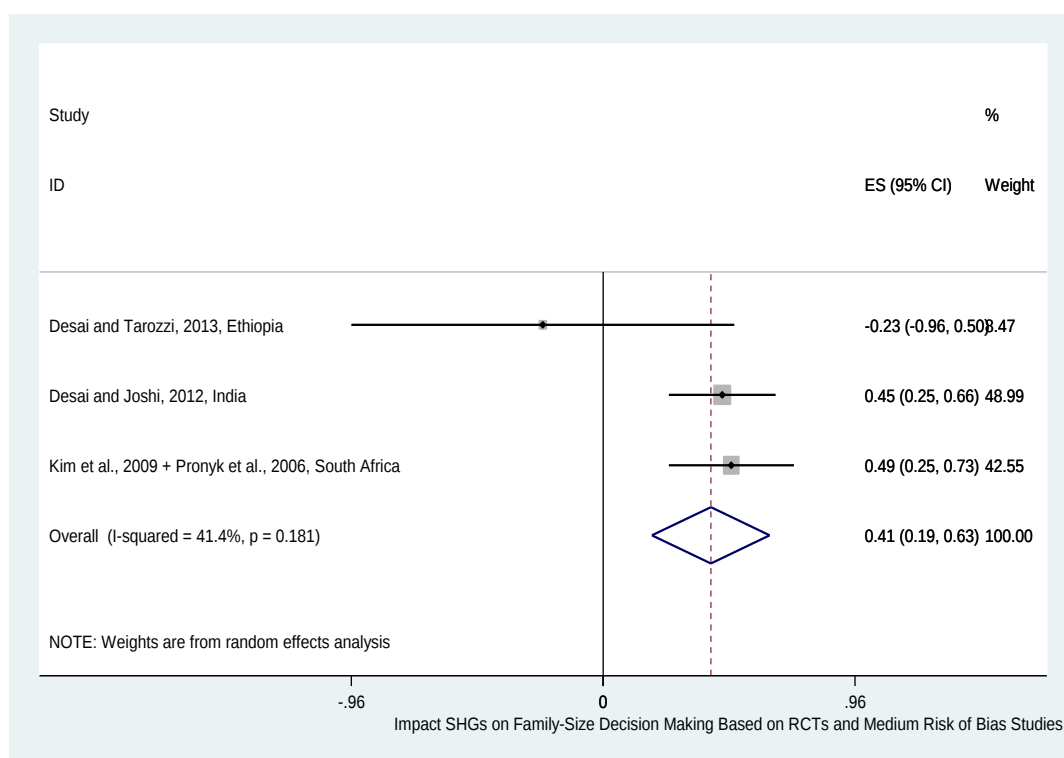
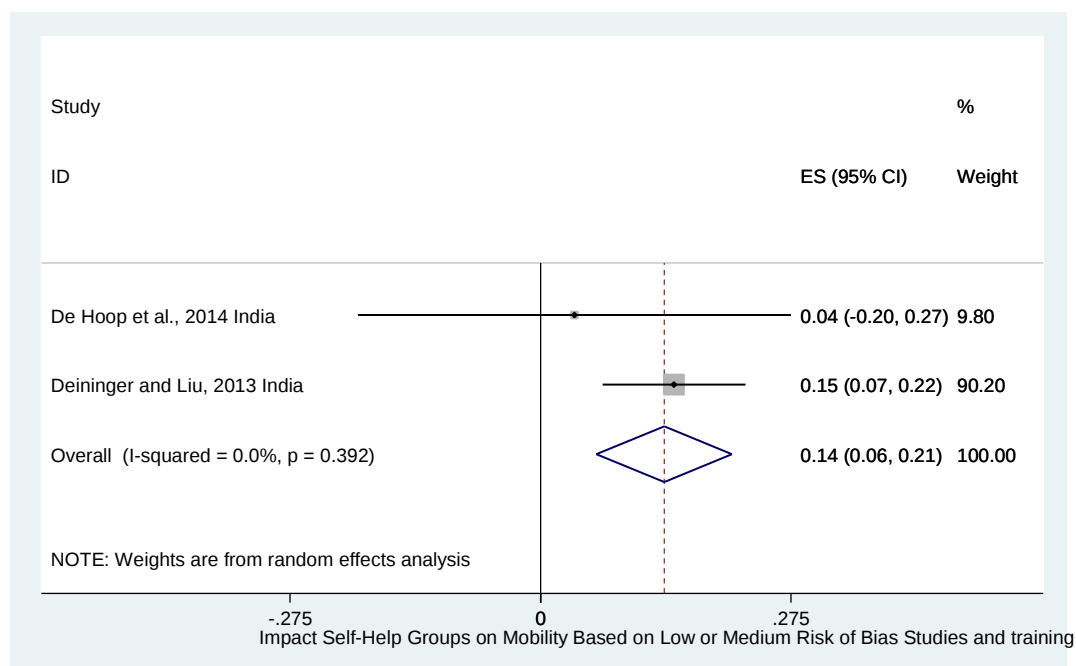


Figure 11. 9: Meta-Analysis for determining the effect of women's self-help groups on women's mobility based on RCTs and quasi-experimental evaluations with a medium risk of selection-bias that focus on SHGs with a training component



Appendix 12: Additional quotes by theme

Psychological Empowerment

Agentic Voice

Author	Quotation
Kabeer, 2011, Bangladesh	"If I have money, I can meet the needs of the stomach; I can buy a new sari and keep it in stock; I can go into society and speak out holding my head high; I can send my children to school. But if I have no land or money, I cannot speak."
Kumari, 2011, South India	"One of the things I have learned is to be able to speak in front of a group of five people without shivering."
Kilby, 2011, South India	"In one SHG, a member referred to having been 'introverted' from harassment, but as a result of the self-help group programme had become 'bold' and gained her 'voice'."
Dahal, 2014, Nepal	"My confidence level is increasing. Before, I was afraid to speak out what I disliked, but now I am not dependent on anyone and I can speak my thoughts and I don't care whether someone likes it or not"
Ramachandar & Pelto, 2009, South India	"Now I understand how to talk to educated urban people."

Participation in Household Negotiations

Author	Quotation
Kabeer, 2011, Bangladesh	"No man in this village ever made a land deed in their wives' names, but now they are registering deposit savings schemes and insurance policies in their wives' names."
Dahal, 2014, Nepal	"I have realized that my views and comments are helpful in making a decision. If it is a family, decisions must be mutual."
Kumari, 2011, South India	"When children are not well, the wife takes the children to hospital even if the husband is not around."
Mercer, 2002, Tanzania	"Being allowed to have money and decide on how to spend it has brought us development in our household and now husbands give us the freedom to do our own things."
Ramachandar & Pelto, 2009, South India	"After two years, they [husband and in-laws] understood the value of the women's groups and remained silent."

Domestic Disputes

Dahal, 2014, Nepal	"The group members came to my house and dealt with my husband and mother in law. I did not want my husband to get jailed but wanted him to behave properly with me. The counseling of the group has helped me have a normal life back again"
Kabeer, 2011, Bangladesh	"Nowadays husbands in villages don't beat their wives so much. They realize that their wives also work."
Kilby, 2011, South India	"Seeing the women free from violence and ill-treatment at a community level and personal level [that] was the strongest form of accountability"
Kumari, 2011, South India	"You cannot come drunk and batter me, my SHG will question you if you touch me, you should be prepared to answer them"
Ramachandar & Pelto, 2009, South India	"My husband used to beat me for joining the [SHG] and my in-laws insisted that he beat me, but I stayed silent and today he does not dare to touch me."
Kabeer, 2011, Bangladesh	"If a husband is beating the daylights out of his wife, five of us women go there and warn him not to make trouble. Because we took this training for arbitration, we are able to talk like this. I could not have done this earlier."
Mathrani & Parodi, 2006, South India	"If our husbands harass us, we do not feel intimidated as we now have a refuge to which we can take our recourse."
Knowles, 2014, South India	"Women in SHG find less fighting between husbands of SHG members due to influence and allegiance of SHG members ... more harmony in the village ... more unity between women and men because of the SHG"
Sahu & Singh, 2012, South India	"Previously my husband used to shout if I had not cooked on time, but now, he adjusts if some day, I am late due to group meetings".

Improved Networking

Author	Quotation
Knowles, 2014, South India	"SHG members complain if a tap is broken or if there is stagnant water ... they bring this to the panchayat [village leader] president's attention issues in the community ... if they have other difficulties they go to government officials now"
Kabeer, 2011, Bangladesh	"Earlier if I saw a group of people sitting together, I did not have the courage to go up to them and say anything. Now even if there are 100 people sitting together, I can go up to them and have my say. Earlier, if we saw a policeman on the road, we would run away. Now even if we go to court, we can talk to policemen there."

Kilby, 2011, South India	"The women themselves insisted on dealing with the tractor owners directly and 'held out' for three weeks before the tractor owners agreed to deal with the women directly. It was the close interaction with staff at all levels, which gave the women the confidence to deal with higher caste village people in this way."
Kumari, 2011, South India	"I went to the panchayat [village leadership]. They asked me where I was from. I said I belong to [the self-help group]. Immediately the staff was asked to take the record and hand it over to me. A [record] was given to me immediately. It was then that I understood the value of belonging to [the SHG]."

Solidarity

Author	Quotation
Dahal, 2014, Nepal	"Our strength is that we have some common problems which we have to solve together. We were deprived of our rights and respect for years and this agony has helped us to move together and form a unity."
Kabeer, 2011, Bangladesh	"One stick can be broken, a bundle of sticks cannot. It is not possible to achieve anything on one's own. You have no value on your own. Now if I am ill, my [SHG] members will look after me."
Kumari, 2011, South India	"Women now have the courage to address [unfair] matters because they say, 'I am not alone. The group members are behind me.'"
Mathrani & Parodi, 2006, South India	"If we disapprove of something, we are able to express our opinions to the larger community as we have a collective voice."

Community Respect

Author	Quotation
Dahal, 2014, Nepal	"The society's view upon being a SHG member has changed. Before it was against the social norms to go out of a house but now society praises women who are involved in SHGs"
Kabeer, 2011, Bangladesh	"There is no proper treatment or medicine in hospitals. We have demonstrated in [our] town, demanding our rights and protesting against the corruption of doctors and theft of public medicine. So now when they hear at the hospital that someone is from our [SHG], they give them a bit more respect."
Kumari, 2011, South India	"When people know we are from GSGSK, we are given special consideration. They give us a chair to sit wherever we go."

Ramachandar & Pelto, 2009, South India	"The biggest benefit of the [SHG] is that we get prestige and honour in our community; we gain experience going to the bank and meeting with officials."
--	--

Financial Skills

Author	Quotation
Kumari, 2011, South India	"The fear of handling money is gone."
Mercer, 2002, Tanzania	"Being allowed to have money and decide on how to spend it has brought us development in our household and now husbands give us the freedom to do our own things."
Knowles, 2014, South India	"Women can go to the bank now without husbands."
Ramachandar & Pelto, 2009, South India	"I handled all the money matters, including buying and selling of chickens and meat."
Macleane, 2012, Bolivia	"The interest rate is really high. Don Pedro—my husband—tells me off: 'Why are you just working for that [the credit]. You're just working for the bank, and the interest is really expensive!'"
Mathrani & Parodi, 2006, South India	"What kind of structure have these women constructed? They are like monkeys, if we hit their home it will collapse."

Catalyzing Broader Social Action

Author	Quotation
Knowles, 2014, South India	"SHG members [have] become councillors, government officials ... those elected [in] six out of 15 wards are women and members of elected panchayat bodies. They advanced their skills and were respected by the community."
Sahu & Singh, 2012, South India	"In the previous election, the MLA candidate had promised to build a road but he did not. When he came for campaigning this time, we questioned him for not keeping his promise and we didn't vote him either."

Understanding Political Context

Author	Quotation
Pattenden, 2011, South India	"A group from another village who had approached the GP [Gram Panchayat—local government] building to request the disbursement of anti-poverty resources had been stoned."

Kumari, 2011, South India	"Empowerment? There has not been complete empowerment. More factors are needed like equal wages. I would say that only 5 to 10% of empowerment has happened."
---------------------------	---

Adverse Outcomes

Barriers to Participation

Author	Quotation
Dahal, 2014, South India	"The issue of selection bias can be agreed to a certain extent acknowledging to the fact that very poor people cannot afford the membership fee and enough time for group activities."
Mercer, 2002, Tanzania	"Some women don't join because they feel inferior, they think that members are rich, can afford things and can be close to the Church, they are in good positions."
Mathrani & Parodi, 2006, South India	"The larger community was of the view that sangha formation is relevant only for the lower castes and that women from the upper castes were demeaning themselves by getting involved in this work."

Disappointment

Author	Quotation
Maclean, 2012, Bolivia	"SHGs operate at very low cost, have a small fund, raise little interest so we cannot accomplish bigger projects and this is our weakness."
Mercer, 2002 Tanzania	"Other women are discouraged because it is almost four to five years since we contributed the money for the cows and up to now we haven't seen any good profit."

Mistrust and Corruption

Author	Quotation
Maclean, 2012, Bolivia	"I don't like [to be treasurer]. It's dangerous. The money can disappear, you can get confused. Even Dona Feliza [a younger woman who was educated in la Paz] can get a little confused sometimes. And they talk about the treasurer and accuse her of things."
Dahal, 2014, South India	"Accounts are not maintained. The leaders of SHGs are heard to have lent the saved amounts to others at high interest rates for personal benefit."

Stigma

Ramachandar & Peltó, 2009, South India	The men used to make comments such as, these women are doing "tamasha" (showing off) and they are going to close down our sangha after a few days. But we did not worry about those comments."
Mathrani & Parodi, 2006, South India	"They think women are attending meetings to get money and take control of the village council." "Men say that women are being overly ambitious." "Upper castes say, 'These women attend meetings and visit the panchayat to get money. They are trying to usurp the position of the gowda and take control of the village.'"

8. Contribution of authors

The study was led by Carinne Brody (CB). This report was written by CB and Thomas de Hoop (TH). CB, Megan Dunbar (MD), Padmini Murthy (PM) and Shari Dworkin (SD) authored by study protocol. The search strategy was developed by CB, MD and Tara Horvath, a search specialist from the University of California, San Francisco. CB and RW, along with several research assistants, conducted the search. The meta-analysis was undertaken by TH and Martina Vojtkova. The qualitative synthesis was undertaken by CB and RW. RW edited the report. CB and TH will be responsible for updating this review as additional evidence accumulates and as funding becomes available.

9. Declarations of interest

The authors declare that we are not aware of any conflicts of interests. Thomas de Hoop was involved in a primary study on the effects of women self-help groups on women's empowerment in Odisha, India. In the risk of bias assessment for this study, we emphasized the opinion of the other reviewer of the quantitative risk of bias assessment Martina Vojtkova rather than relying on the opinion of Thomas de Hoop.

10. Sources of support

External sources

Funder: International Initiative for Impact Evaluation

International Initiative for Impact Evaluation
London International Development Centre
36 Gordon Square
London WC1H 0PD
United Kingdom

3ieuk@3ieimpact.org
Tel: +44 207 958 8351/8350



www.3ieimpact.org