ECWG
EVIDENCE CONSORTIUM ON WOMEN'S GROUPS

ECWG WORKING PAPER 02 NOVEMBER 2020

Economies of Scale of Large-Scale International Development Interventions: Evidence from Self-Help Groups in India

Garima Siwach*, American Institutes for Research
Sohini Paul, Population Council
Thomas de Hoop, American Institutes for Research

The authors would like to thank the Bill & Melinda Gates Foundation for supporting this study. We would especially like to thank Shubha Jayaram, Aishwarya Lakshmi Ratan, and Subhalakshmi Nandi at the Gates Foundation, David Seidenfeld from the American Institutes for Research, and Sapna Desai from the Population Council for their valuable comments and feedback on this paper. Finally, we would like to extend our gratitude to the staff at *Jeevika* – the Bihar Rural Livelihoods Promotion Society – for their advice and engagement.

*Corresponding author. Email address: gsiwach@air.org

POPULATION COUNCIL Ideas. Evidence. Impact. Stanford University EVANS SCHOOL OF PUBLIC POLICY & GOVERNANCE UNIVERSITY of WASHINGTON Evans School Policy Analysis and Research (EPAR) theAfricaCentre MAKERERE UNIVERSITY COLLEGE of HEALTH SCIENCES MAKERERE UNIVERSITY School of Women and Gender Studies AIR AMERICAN INSTITUTES FOR RESEARCH Campbell Collaboration

Contents

Abstract

Recent studies have suggested that development programs, including those that focus on microfinance, financial literacy, and women's empowerment, often show reduced impacts when scaled up. Yet, scaling up may also lead to economies of scale, which can reduce marginal costs of operations and maintain the program cost-effectiveness despite lower impact. This paper shows evidence on the association between program scale, costs, and cost-effectiveness of large-scale development interventions by analyzing how the costs of a large-scale Self-Help Group (SHG) program in India changed over a ten-year period. The study focuses on program costs of *Jeevika* – the Bihar Rural Livelihoods Promotion Society – over eight years from 2007 to 2015. Using expenditure data from program's annual audit statements, we estimate costs of different program components, including capacity building, institutional development, community investment, and project management. Our findings show that per capita expenditure on basic program activities declined from $37 in the program's first year when it covered approximately 8,000 members to $13 after eight years when it reached approximately 5.7 million members. Program per capita expenditure declined by $3 for every 100,000 additional households, but the decline slowed down with a large jump in scale. Combining these cost estimates with existing estimates on program impact from the pilot phase and the scaled-up phase, we found that the program led to reductions in household debt from high cost loans by $85 in the initial phase and by $33 in the scaled-up phase, but we found no meaningful difference in Cost-Effectiveness Ratios between the two phases because of lower costs caused by economies of scale.

Introduction

In recent years, the developing world has seen an increased focus on institutionalizing women's groups with economic objectives as a key channel to improve women's empowerment and economic outcomes. These interventions include self-help groups (SHGs), savings groups, mothers' groups, health groups and community mobilization groups (Diaz-Martin et al., 2020). While interventions that involve women's groups have a long history in community development, these networks have traditionally been more informal and disaggregated than they are envisioned today. Many countries are moving towards large scale investments to mobilize women in group settings and use these groups as delivery vehicles for various forms of economic and health initiatives. For example, India's National Rural Livelihoods Mission (NRLM) – the largest poverty alleviation initiative in the world – aims to mobilize 70 million households into SHGs and provide them with means to sustainable livelihoods enhancements and access to financial services (Ministry of Rural Development, 2011). Similarly, in Nigeria, the Nigeria for Women Project recently began mobilizing 0.32 million women into Women's Affinity Groups to build women's social capital and introduce them to livelihoods programs in six states of the country (World Bank, 2018a).

To some extent, these decisions to scale up can be justified by evidence on the impact of women's groups. For example, a systematic review showed that SHGs can have positive effects on several domains of women's empowerment—including economic, reproductive, social, and political domains—although impacts depend substantially on program design and implementation context (Brody et al., 2015). In addition to the impact of economic SHGs, research has also shown that open, community groups of women that involve participatory learning outcomes can improve maternal and newborn health in low-resource settings (Prost et al., 2013). Some studies attribute these greater benefits to accumulation of social capital and mutual accountability resulting from peer interactions and social exchanges facilitated by groups (Gram et al., 2019; Brody et al., 2015; Desai et al., Forthcoming). However, a recent evidence synthesis by Diaz-Martin et al. (2020) suggests that the benefits of women's group programs primarily result from leveraging groups as a platform to deliver programming to many women at once, suggesting that groups may be able to deliver benefits at a low cost per program participant. Based on this finding, the authors argue that more research is needed to understand the overall costs of group models, including costs of creating and mobilizing groups, and the operational costs of group-based activities (Diaz-Martin et al., 2020).

While various studies examine the costs and cost-effectiveness of women's groups aiming to improve health outcomes (Mangham-Jefferies et al., 2013; Colbourn et al., 2015; Sinha et al., 2017; Pulkki-Brännström et al., 2020), somewhat paradoxically there is limited information on costs of women's groups with economic objectives, like SHGs or savings groups. Studies that do include some estimation of cost-effectiveness of such programs either use aggregate program funding amounts (Deininger & Liu, 2015), or focus on costs of small-scale programs (Venton et al., 2013). Given the growing interest in expanding the outreach of these programs however, it is imperative to study not only the costs and outcomes of small pilot programs, but also the potential role of scale in influencing program impact, implementation, and costs.

Program scale may impact outcomes through different mechanisms. On the one hand, increase in scale may lead to an adverse impact on implementation quality leading to lower direct effects on beneficiary outcomes. A recent meta-analysis of development programs, including those that focus on microfinance, financial literacy and women's empowerment, concluded that studies that had larger participant size as well as programs that were implemented by governments had lower effects (Vivalt, 2020). Indeed, program-specific evidence on other development interventions have shown that many pilot programs that are successful at a smaller scale fail to achieve similar positive effects after scaling up (Bold et al., 2018; Bryan, Chowdhury and Mobarak, 2014). On the other hand, service scale may also alter the costs of service provision if programs are able to achieve economies of scale (Tulloch, 2019). The scale-up of SHG programs and other development interventions could thus either increase or decrease their cost-effectiveness depending on how scale changes the implementation quality, effectiveness, and costs of programming. In the absence of cost information, and specifically cost information about programs delivered at different levels of scale, we are severely limited in our ability to understand the overall returns on investment of large-scale programs and interventions.

This paper contributes to the literature by examining how the scale of an SHG program in India influences program costs, and potentially, program cost-effectiveness. We study how the costs of a large-scale SHG program in the state of Bihar – *Jeevika* – changed over time since its inception in 2007 when it mobilized 8,000 women into SHGs, to its current reach where it has mobilized over 9 million women into 0.8 million SHGs. *Jeevika* has been operating the Bihar Rural Livelihoods Project (BRLP) since 2007, and currently implements India's NRLM through Bihar's State Rural Livelihoods Mission (SRLM). The NRLM, which was launched in 2011, operates in 28 states through the SRLMs, which create and work with women's SHGs to facilitate institutional and capacity building, financial inclusion, livelihoods promotion, social

inclusion, and development. SHGs under NRLM (and BRLP) usually start with a period of collective savings to facilitate intragroup lending, after which members can gradually take larger loans (usually from formal financial institutions like banks). In addition, many SHGs include additional training components to improve agricultural and non-agricultural livelihoods and health outcomes, including training for some of the most disadvantaged groups, including women from scheduled castes and scheduled tribes. Using program expenditure data from annual audit reports of *Jeevika* between 2007-08 and 2015-16, we estimate costs of different program components, including capacity building, institutional development, community investment, and project management. We then use these estimated costs to empirically test for evidence of economies of scale. In addition, we provide some evidence on how costs of the *Jeevika* program change with the inclusion of extra programming, and how program cost-effectiveness may change with respect to scale.

Understanding how costs change with respect to scale is challenging since the nature of program and the number of activities often change as programs expand. Researchers have pointed out that the ideal way to differentiate between short-run and long-run costs is by exploiting time-series data on program costs – a challenging endeavor given the severe lack of data on costs over time (Johns and Terra, 2005). Instead, most studies observe cost data at one point in time precluding researchers to study how costs vary with changes in both program scale and program scope (in terms of the number of outputs produced). These factors hold implications for women's economic groups like SHGs, that usually start with the goal of increased financial access to women, but with time, often include additional layers like health or farm and non-farm-based livelihoods initiatives.

Prior costing analyses and systematic reviews of SHG programs also point to a trade-off between cost savings and benefits from increasing the scope of services. Evidence suggests that both the benefits and the costs of SHGs increase when an increased number of activities are delivered. For example, various studies show wide variation in SHG program implementation costs by range of activities and implementation organizations (e.g. Isern et al., 2007; Tankha, 2002; Harper, 2002). SHGs that focus on empowerment, social change, and livelihood generation have higher costs than SHGs that focus solely on lending. At the same time, Brody et al. (2015) showed that SHGs' impact on women's empowerment increased significantly when they included a training component with a focus on business skills, community development, or employment and leadership. In the absence of reliable cost data, it thus remains unclear whether adding in extra programming to group-based models with

economic objectives will increase the cost-effectiveness of women's groups (Diaz-Martin et al., 2020).

So far, researchers have only examined how the effectiveness and implementation of SHGs and specifically the *Jeevika* program varies with scale of programming (Hoffmann et al., 2020; Majumdar, Rao, & Sanyal, 2017). Findings from impact and process evaluations of *Jeevika* suggest that the impacts of the program on women's empowerment and economic outcomes reduced after expansion in the pace and scale of implementation because of a lighter program touch (Hoffman et al., 2020). Specifically, researchers found that the program was less consistent in applying some of the ritualized processes of group activities[1] after the scale-up, which likely reduced the effectiveness of the program (Majumdar, Rao, & Sanyal, 2017; Kochar et al., 2020). Other studies suggest that scaled-up programs may have larger general equilibrium effects [i.e. through changes in prices and wages (e.g. Breza & Kinnan, 2018; Muralidharan, Niehaus, & Sukhtankar, 2020)]. These general equilibrium effects may, however, increase even further when improved implementation quality results in increases in program participation. A cluster-randomized controlled trial of *Jeevika* suggests that the program led to changes in community-level prices and particularly a reduction in interest rates charged by informal money lenders, likely due to reduced demand for informal credit (Hoffman et al., 2020). A follow-up study by Kochar et al. (2020) showed that reductions in high-cost loans (mostly because of changed source of borrowing) remained larger for a treatment group that participated in the program for 7 years than for a control group that participated in the program for four years. The program may result in even larger reductions in interest rates charged by informal money lenders when improved implementation quality would result in improved access to formal credit for SHG members.

Jeevika makes a compelling case for examining how scale influences program costs. The program originally started in six priority districts of Bihar with funding from the World Bank in 2006-07, before NRLM was formally launched by the Government of India. In 2011, with additional funding from the World Bank as well as the launch of NRLM, *Jeevika* scaled up its operations beyond the originally planned six districts. In our data, we observe program expenditure over the short-run when the program operated at smaller scale, and in the longer-run when the program scaled up across the entire state. These time series cost data enable us to distinguish between short-run and long-term costs. In addition, the data allow for

[1] These processes were originally termed as the *Panchasutra* that included five activities – regular meetings, regular savings, regular inter-loaning, timely repayments, and keeping the book of accounts up to date.

differentiating between cost components related to different program outputs, thus studying both changes in economies of scale, as well as economies of scope.

Further, we examine how the cost-effectiveness of the *Jeevika* program could potentially change with respect to scale by comparing a cost-effectiveness ratio (CER) for household benefits from reduced dependence on high cost loans after an initial pilot of the program in its early years with an estimate of the CER after the scale up of the program. We selected reduced dependence on high cost loans as the main outcome of interest because previous evaluations of the *Jeevika* program found the most direct effects on this outcome (Hoffman et al., 2020; Datta et al., 2015). Although this study does not estimate a complete return on investment of the program based on a holistic range of outcomes, a cost-effectiveness estimate is helpful in analyzing the cost of an additional unit of impact measured for the same outcome during the two phases of the program.[2]

Our findings show evidence for economies of scale, demonstrating the importance of not only examining the relationship between scale and the effects of international development programs but also the relationship between scale and costs of such programs. Specifically, we find that the annual per-household expenditure for basic program activities under *Jeevika* declined from approximately $62 per member at the start of the program in 2007, when the program served fewer than 10,000 households, to almost $12 per member at its scaled-up level in 2015-16, when the program reached more than 5 million households. We also substantiate our findings on cost efficiencies with respect to scale and demonstrate external validity by exploiting time series data on program expenditure from three other programs involving BRAC's women's groups focusing on microfinance across diverse settings in Bangladesh, Uganda, and Tanzania. Similar to our main findings, operating costs of BRAC microfinance program reduced significantly with respect to program outreach.

A comparison between the CER of the *Jeevika* program in terms of cost per dollar of decline in high cost loans after the program scale up and the CER after it was initially piloted suggests that the scaled-up version of *Jeevika* may be as cost-effective in achieving benefits from reduced dependence on high cost informal loans as its pilot version because of economies of scale. Despite its smaller impacts on women's empowerment and income, a combination of impact

[2] An updated impact evaluation of *Jeevika* looking at longer-term effects was completed in 2020 – see Kochar et al. (2020). In collaboration with the study team of the impact evaluation, we are currently pursuing a complete cost-benefit analysis of the program that expands the current analysis to multiple outcomes (Paul et al., forthcoming).

and simulation analyses shows that the CER for interest savings of the scaled-up *Jeevika* program is similar to the pilot version of the *Jeevika* program.

Background & Program Description

Located in the Northern belt of India, Bihar is a primarily rural state, with 89% of its population residing in rural areas (Government of India, 2011) and has historically performed lower than other states on most social and economic development indicators. In 2012, the second largest share of the country's poor lived in Bihar, which had a poverty rate of 34%, significantly higher than the national average of 22% (World Bank, 2018). In addition, Bihar grapples with extreme gender disparities in multiple domains. At just 9% and 15% respectively, the state has the lowest female labor force participation rate and the second lowest rate of secondary education attainment among women in the country (World Bank, 2018). Data from the National Family Health Survey-4 also indicate that women in Bihar had much lower indicators on financial inclusion. In 2016, 33% of women in Bihar reported having "money that they can decide how to use" and 26% indicated having bank or savings account that they themselves use, compared to a national average of 42% and 53% respectively (IIPS and ICF, 2017).

In 2005-06, with support from the World Bank, the Government of Bihar initiated planning for the launch of the Bihar Rural Livelihoods Project (BRLP). The structure and goals of the BRLP were defined along similar poverty alleviation projects in other states,[3] with an initial goal of improving livelihoods of the rural poor by developing institutional capacity through women's SHGs and linking them to formal financial institutions and other agencies to negotiate better services. In 2007, the Government of Bihar established the Bihar Rural Livelihoods Promotion Society, also known as *Jeevika*, in six priority districts of the state with the aim of mobilizing poor households into SHGs under the BRLP. In 2011, the Government of India formally launched the NRLM to promote institutions of the rural poor including social collectives and livelihoods institutions across the country, using women's SHGs as the primary platform. As a result of the NRLM launch and with additional funding from the World Bank in 2011–12, *Jeevika* expanded its services and the scale of SHGs to cover a greater number of districts. At its inception, *Jeevika* received $63 million from the World Bank and an additional $7 million from the State Government of Bihar (World Bank, 2007). The organization received additional funding from various implementing partners—such as the United Nations Children's Fund (UNICEF)—for

[3] The BRLP specifically borrowed from the model of the Society for Elimination of Rural Poverty (SERP) in Andhra Pradesh.

specific interventions, totaling approximately $3 million initially (*Jeevika*, n.d.). Under the NRLM—that is, after 2011— *Jeevika* was jointly financed by the State Government as well as the central Indian government, which sourced additional funding from the World Bank. In 2012, the program received another $100 million from the World Bank. These funds were used to scale up core program interventions to all blocks in the six initially chosen project districts and to scale up interventions to new districts, thus deepening livelihoods interventions and enabling convergence with other development programs—such as the Mahatma Gandhi National Rural Employment Guarantee Scheme—and strengthening partnerships and monitoring and evaluation systems (World Bank, 2012). After its initial pilot in six districts in 2007 the *Jeevika* program quickly scaled up to reach more than five million households by 2015-16 and had mobilized over nine million households by 2020.

SHGs under *Jeevika* follow a federated structure involving a system of functional integration between groups at various levels (Shylendra 2018). Federated SHGs are typically organized in a four-tier pyramidal structure. The basic unit, SHGs, are federated to primary level federations (PLFs), which are federated further into secondary level federations (SLFs). Finally, SLFs are federated to apex level federations (ALFs). *Jeevika* follows a three-tier structure where the basic unit is the SHG. An average *Jeevika* SHG consists of 10 to 15 women who come together as a platform for accessing program services that include financial inclusion in form of low-cost credit and access to formal banking, as well as basic literacy and livelihoods training (Hoffman et al., 2020). In the second tier, SHGs are federated into PLFs known as Village Organizations (VOs), which are further federated into Cluster Level Federations which serve as the ALFs. VOs and CLFs facilitate collective action, adoption of livelihoods enhancement and income generating activities, and development of linkages with market institutions (World Bank, 2017).

Program expenditures under *Jeevika* are reported for four broad components: Community Institution Development (CID), Community Investment Fund (CIF), Special Technical Assistance Fund (STAF) and Project Management (PM). The CID fund provides all expenses for institution building, including the costs of mobilizing groups, and capacity building of staff at the district, block, and cluster level. The CIF component provides direct grants to the community through the Revolving Funds (RFs), as well as through Vulnerability Reduction Funds (VRFs). The RF aims to meet the credit need of group members and is considered “catalytic capital” for leveraging repeat bank finance, thus promoting income generating activities, for example through investments in livelihood promotion and skills development. VRF is usually provided to VOs to address risks, including food security risks, health risks, and natural disasters through

the Food Security Fund (FSF) or Health Security Fund (HSF). The STAF provides technical assistance to the formal financial sector to support microfinance initiatives and pro-poor banking and promoted public-private partnerships to improve the quality and quantity of public services delivered to communities. STAF resources are also used to build a partnership with other stakeholders and to organize training programs for staff to strengthen the service delivery mechanism. Overall coordination, project implementation, financial management and monitoring and evaluation (M&E) at the state and district levels are supported through resources under the Project Management (PM) component.

Data

***Jeevika* Program and Cost Data**. We obtained data on program implementation, components, outreach and expenditures from *Jeevika* Annual Reports and audited financial statements from 2007-08 to 2015-16. These data provide information on annual program outreach and scale, including the number of households mobilized into SHGs, number of SHGs promoted, number of VOs and CLFs, number of SHGs linked to a bank account, number of households involved in farm and non-farm based livelihoods, and other program outputs like the number of VOs that implemented health and food security interventions. Data on program expenditures were collected from annual audit statements of *Jeevika* and World Bank project documents. All expenses were categorized under the following: 1) CID which included social mobilization and institutional building; 2) CIF which included community investments in livelihoods; 3) Project management; and 4) STAF expenditures.

It is important to note that our analyses are only based on reported program expenditures. Ideally, a costing analysis should also include costs of resources that were utilized but not directly paid for – for example, the time spent by volunteer staff, or use of buildings and resources for BRLP or NRLM-specific activities that may have been already in use and paid for by another government entity. All costs presented in this study should therefore be considered as an underestimate of true program costs and should be regarded as program accounting costs only. For this reason, we should also exercise some caution in the interpretation of the cost-effectiveness estimates.[4,5]

[4] The rest of this paper uses the terms “costs” and “expenditures” interchangeably.
[5] We are planning to conduct a full holistic cost-effectiveness analysis of the *Jeevika* program through the collection of primary cost data but had to postpone these plans because of the emergence of COVID-19.

***Jeevika* Outcomes and Impact Data.** We use findings from impact evaluations of the *Jeevika* program to complement our costing analysis and generate evidence on cost-effectiveness. For this paper, we consider the most direct source of program impact, in the form of cost savings from borrowing at lower interest rates. Financial inclusion, including access to and use of formal banking services, is the first area of program focus after group mobilization. Findings from two impact evaluations (Datta et al. 2015; Hoffman et al., 2020) – the first focusing on the first phase of the project (pre-2012) and the latter on the second phase (2012-2014) showed that the program resulted in a statistically significant and sizeable reduction in interest rates on loans taken by households, primarily due to reduced dependence on informal moneylenders.[6] The extent of aggregate benefits from these cost savings depends on how likely program participants and non-member households in the same Gram Panchayat were to borrow at lower interest rates. To analyze this question, we accessed and analyzed data from the two impact evaluations of the *Jeevika* program posted on Harvard Dataverse (Datta & Rao, 2018a; Datta & Rao, 2018b) and complemented the analysis with published literature (Hoffman et al., 2020; Datta et al., 2015). The phase 1 survey dataset included 3,997 observations, 50% of whom belonged to treatment areas where *Jeevika* entered at various points during 2008. The phase 2 dataset included 5,870 observations, 50% of whom belonged to treatment areas where *Jeevika* entered between January and April 2012.

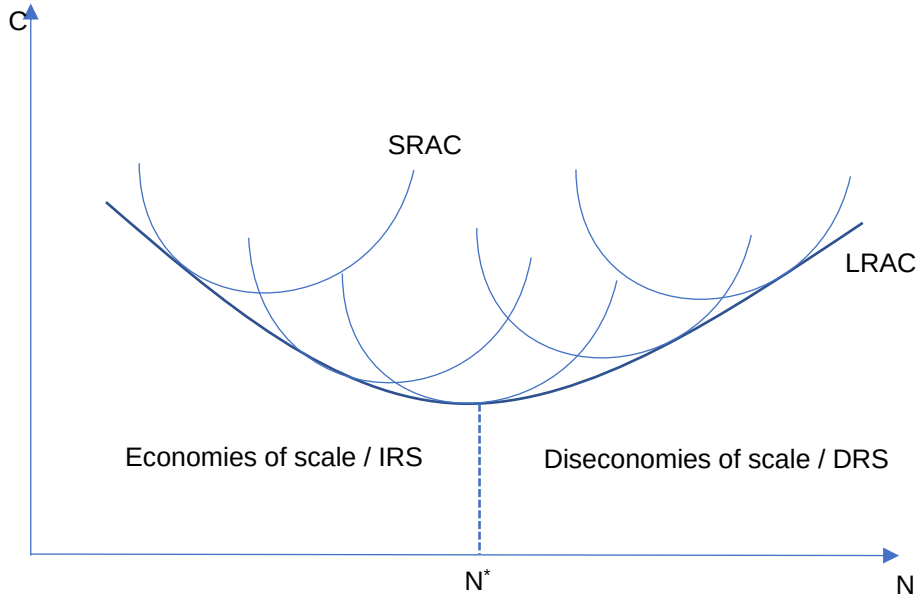
Theoretical Framework

We present a simple theoretical framework to generate hypotheses on the relationship between the scale-up of *Jeevika*, and its costs, effects and cost-effectiveness, starting with a production function approach. We define scale as the number of women mobilized under *Jeevika* (Y), which is a function of various inputs, including capital (K), and labor (L). Capital includes other goods like buildings and equipment which are used in the production process. Labor includes managerial staff members who are salaried employees working at the state, district, and block level; and temporary workers including frontline workers such as area coordinators, community coordinators, facilitators, and community mobilizers. Program scale can then be expressed by $Y(K, L)$. Similarly, program costs depend upon the resources utilized in producing a given level of output and can be expressed by $C(Y(K, L))$.

[6] Kochar et al. (2020) conducted a follow-up study to determine longer-term effects of the *Jeevika* program. We will discuss these results in a full holistic cost-effectiveness analysis of the *Jeevika* program.

Following standard microeconomic theory, *Jeevika* can experience different economies of scale based on the association between cost and output (i.e. the number of women mobilized). The program will have economies of scale if average costs reduce with the number of women mobilized; and diseconomies of scale when average costs increase with the number of women mobilized. We present these different scenarios through long-run average cost (LRAC) curves in Figure 1. As Figure 1 shows, program average costs decline with respect to scale once production becomes efficient due to factors like specialization and experience.

Figure 1: Long run cost and economies of scale



For a standard firm, optimal level of scale will be determined based on profit maximization, which corresponds to marginal costs equaling marginal revenue. In our case, we can proxy for *Jeevika*'s objectives by differentiating between the number of women mobilized (output) and the number of women empowered (outcome) from the *Jeevika* program. We assume that the primary objective of the program is to maximize women's social and economic empowerment, which we denote by E, at the lowest cost; or in other words, the program will aim to minimize operational costs for each level of outcome E. The extent to which the program achieves this objective depends, among other factors, on program scale. On the one hand, a higher number of women mobilized implies higher exposure to program and higher number of potentially empowered women. On the other hand, increasing the number of women mobilized may reduce the average probability of achieving the outcome if program scale is negatively associated with program impact. The latter point is, in some part, grounded in empirical evidence.

Implementation research on *Jeevika* found substantial differences in the quality of implementation before 2012 when the program operated at a smaller scale, and after 2012 when the program scaled up across the state (Majumdar, Rao, & Sanyal, 2017).

Based on this empirical background and *Jeevika*'s goal of maximizing empowerment outcomes for the lowest cost, our theoretical framework makes the following assumptions:

(a) The quality of labor involved with the implementation at the local level may be adversely affected as the program scales up (Kochar et al., 2020). Specifically, Majumdar, Rao, & Sanyal (2017) found that compared to the first phase of the program, community facilitators in the scaled-up second phase lacked experience, having spent little to no time in institutional learning. An impact evaluation of the second phase of the program also found lower effects on women's empowerment compared to the first phase (Hoffman et al., 2020). Denoting labor quality by Q, this assumption implies that $Q'(Y) < 0$.

(b) Average costs decline with respect to scale, not only owing to economies of scale, but also due to possible declining labor quality assuming that labor costs are lower for lower quality labor. This implies that $C'(Y) < 0$.

(c) Program impact is a direct increasing function of labor/ program personnel and implementation quality: $E'(Q) > 0$. Combining this with assumptions (a) and (b):

$$Q'(Y) < 0; E'(Q) > 0 \Rightarrow E'(Y) < 0 \quad (1)$$

Jeevika's objective is to minimize the dollar amount spent per unit of impact, which is denoted by the Cost-Effectiveness Ratio (CER):

$$CER = \frac{C(Y(L), Q)}{E(Q(Y))} \quad (2)$$

Change in CER with respect to Y is then given by:

$$\frac{E.C'(Y) - C.E'(Q)Q'(Y)}{E^2} \quad (3)$$

Since $C'(Y) < 0$, $E'(Q) > 0$, and $E'(Y) < 0$, it is *a priori* unclear how overall cost-effectiveness will change with respect to program scale. While economies of scale suggest that costs may decline with increased scale, the change in CER will also depend upon the relative magnitude of

changes in empowerment and changes in costs. In the analysis ahead, we shed some light on this question and show the implications of incorporating scale in questions about costs and cost-effectiveness.

Methods

We examine the costs of increasing both the scope as well as the scale of services with respect to expanding *Jeevika*'s outreach of the SHG program. In terms of program scale, we consider costs from expanding the geographic supply of the program by adding more districts over time. In terms of scope, we study component-specific costs of the program, differentiating between expenditures related to basic program activities (community mobilization and financial inclusion), and expenditures related to providing additional services like livelihoods initiatives, health, and food security.

We estimate average program costs as costs per SHG woman targeted under the *Jeevika* program, treating each *Jeevika* member added as an additional program output. While not all women received the same level of services with the same intensity, we estimate an average cost equivalent to an "Intent to Treat" cost – that is, expenditure per woman SHG member with access to potential program services.[7] From *Jeevika*'s expenditure components discussed earlier, we differentiate between program operating expenses required to cover – 1) primary program activities including community mobilization and financial inclusion (reported under CID expenditure); 2) program technical assistance (reported under STAF expenditure); 3) program management (reported under PM expenditure); and 4) community investment activities (reported under CIF expenditure).

While the first three of these four activities are expected to be carried out across all target areas, not all SHGs may be eligible to receive CIF every year. Specifically, community investment funds are disbursed through village organizations (VOs) and subsequently through cluster-level federations (CLFs), and new SHGs may take time to form these higher-level federations. In addition, *Jeevika* aims to form SHGs that will eventually become self-reliant and self-managed by gradually adopting income generating activities and profits for the federations. In 2017, the World Bank conducted an implementation assessment of the program, and through a random stratified sample, found that 89% of older SHGs (formed prior to 2012) had become self-reliant by 2017 (World Bank, 2017). Although we are unable to obtain precise estimates on SHGs that

[7] This also implies that program attrition is not considered when estimating annual per member costs.

were eligible to receive *Jeevika*'s support on different activities in every year, we estimate three average cost models based on assumptions on the size of the target groups for these activities:

Model 1 – Cumulative Participation: Under this model, we assume that every SHG member under *Jeevika* is eligible for all services, irrespective of when they joined the program. We divide all expenditure components (CID, CIF, STAF, and PM) in each year among cumulative number of women mobilized into SHGs up to that year.

Model 2 – Fixed-Period Participation: Under this model, we assume that *Jeevika* supports all SHGs and related-activities for a period of three years, after which SHGs become self-sustainable. We divide all expenditure components in each year among women who have been a member of *Jeevika* for no longer than three years.

Further, specifically for community investment funds, we estimate one additional model:

Model 3 – VO-based Participation: Under this model, we assume that only women from SHGs that have federated to form VOs are considered as the target outreach of CIF-related activities. This scenario focuses on this sample because the original project plan from 2007 planned to allocate and transfer resources under CIF to VOs as grants (World Bank, 2007). VOs are expected to use this resource as revolving fund to finance various SHGs under the VO based on micro planning processes. VOs are flexible in fixing terms and conditions under which the resources are lent to SHGs, suggesting that different VOs may be using these funds for different activities. Under this assumption, we divide the CIF expenditure each year among women who are members of SHGs that were federated into VOs.

Based on these assumptions, we estimate annual average program costs under each assumption separately for component i in year t as:

$$AC_{it} = \frac{Annual\ expenditure\ on\ component\ i\ in\ year\ t}{Target\ number\ of\ women\ under\ component\ i\ in\ year\ t} \quad (4)$$

Here the target number of women is estimated separately for three models based on the three assumptions defined above.

We extracted all program expenditure data in local currency (Indian Rupee) and converted to 2015 USD by first converting annual expenses to 2015 prices by adjusting for Rupee inflation using the Consumer Price Index (CPI) method, where:

$$Expenditure_{i2015} = Expenditure_{it} \times \frac{CPI_{2015}}{CPI_t} \quad (5)$$

We then converted to 2015 USD by adjusting for 2015 USD market exchange rate.

Finally, we used the production function to test the possibility of increasing, decreasing or constant returns to scale, for the same program over time for each scenario. For this, we estimate the time-varying average cost function as a function of cumulative number of individuals served in first linear, and then quadratic form as shown in equations (6) and (7):

$$AC_t = \alpha_0 + \alpha_1 n_t + \epsilon_t \quad (6)$$

$$AC_t = \alpha_0 + \alpha_1 n_t + \alpha_2 n_t^2 + \epsilon_t \quad (7)$$

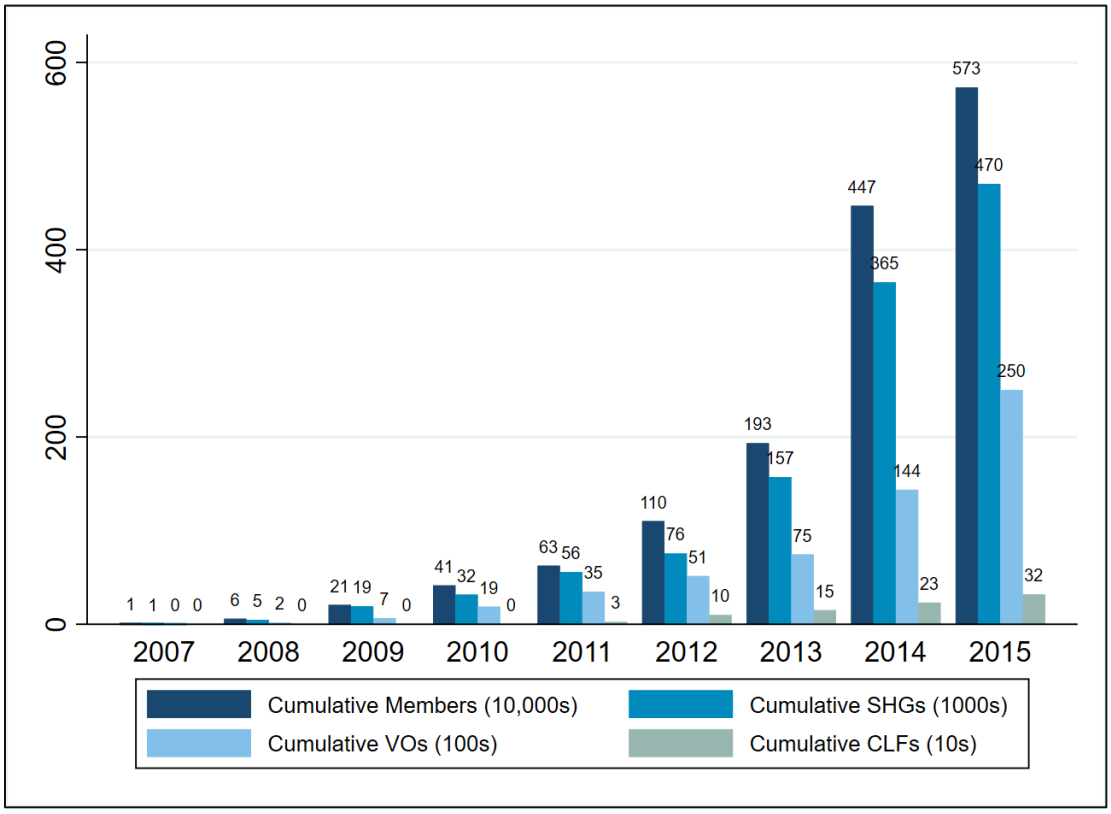
Here AC_t is the average program cost in year t (expressed in 2015 USD), and n_t is the cumulative number of individuals reached by year t. The magnitude of the coefficient α_1 can then be interpreted as the slope of the linear component of the average program cost, and the significance and sign of α_2 will indicate whether the change in costs increases, decreases, or stays constant as the number of women mobilized increases.

Results

Expansion in Scale of Activities

We begin the analysis with an overview of the scale of *Jeevika*'s operations between 2007-08 to 2015-16. Figure 2 shows the number of SHGs, VOs, and CLFs formed by the end of each financial year, indicating a steep increase in mobilization after 2012-13. In 2011-12, *Jeevika* took over the expansion of SHGs under the NRLM and received additional financing from the World Bank for the second phase of the BRLP project. The initial two years of expansion were reported to be relatively challenging because of delays in recruiting adequate staff (World Bank, 2017). However, as shown in Figure 2, the pace of mobilization increased steadily post expansion, especially after 2012-13, and by 2015-16 over 5.7 million women had been mobilized into almost half a million SHGs.

Figure 2. *Jeevika* Program Scale-Up



Non-Community Investment Expenditure

Next, we analyze annual expenditure related to basic SHG programming and program operations. Overall, we find a steep decline in program expenditures per member on basic group activities in initial years of the program, with the decline becoming slower after 2011-12. This change overlapped with the large increase in program scale during the second phase of the project. These expenditures include amount spent on activities like mobilization, group meetings and financial inclusion (CID), project management (PM), and technical assistance (STAF). In addition, post-2013, project expenditure also included separate expenditure categories for operating the State Rural Livelihoods Mission (SRLM) under the NRLM. Figure 3 shows the change in per member expenditure on these activities under two assumptions – 1) assuming that cumulative number of program beneficiaries are eligible for support in a given year (Model 1); and 2) assuming that all beneficiaries are eligible for program support for up to three years from year of initial participation (Model 2).

Figure 3. *Jeevika* Total non-CIF Expenditure Over Time

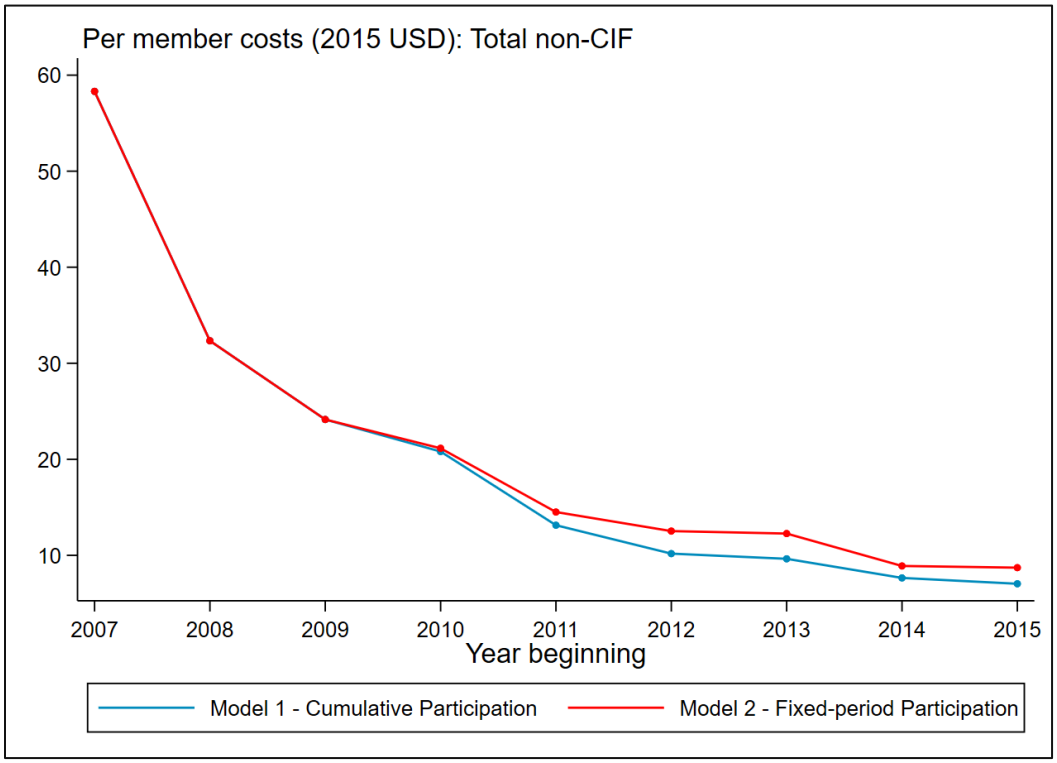


Table 1 presents the results more formally using regression estimates of the cost function. When only considering the linear estimate, results suggest that every 100,000 members increase in outreach is associated with a $0.48 decline in expenditure per capita in Model 1, and $0.56 decline in Model 2. These estimates change significantly when we control for the quadratic outreach term, suggesting that the marginal change in cost decreases with respect to increase in outreach. Specifically, the estimates suggest that expenditure per capita for the smallest level of outreach ranges from $35.6 in Model 1 to $37.4 in Model 2. Beyond that, increasing scale by an additional 100,000 members is associated with a decrease in per capita expenditure by $2.2 to $2.8. The significant positive relationship between outreach squared and per capita expenditure indicates that the decline in costs per capita slows down with a large increase in scale. As scale increases by a large number, existing inputs may reach their threshold of productivity and the program may require recruiting a large amount of new inputs which drives up the operational costs. During the period of analysis, the largest level of program participation was 5.73 million. Our regression estimates suggest that per capita expenditure for this level of outreach was $12.05 under Model 1, and $13.47 under Model 2.

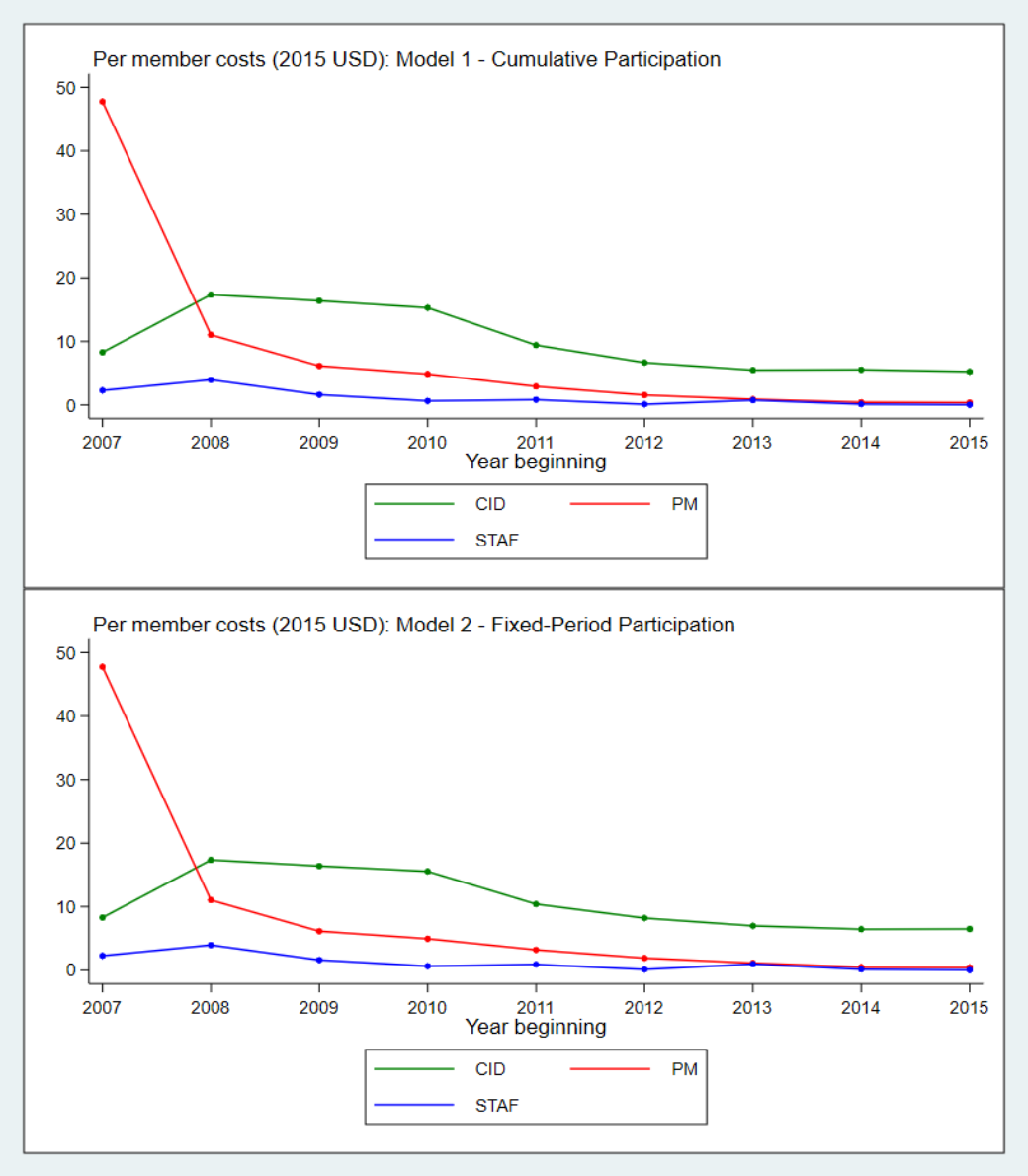
Table 1. *Jeevika* Total non-CIF Expenditure

	Model 1 Cumulative Participation		Model 2 Fixed Period Participation	
Reach (in 100,000)	-0.481*	-2.191**	-0.560*	-2.824**
	(0.240)	(0.892)	(0.279)	(1.095)
Reach (in 100,000) squared		0.031*		0.050*
		(0.016)		(0.024)
Constant	28.135***	35.640***	28.976***	37.447***
	(6.104)	(6.396)	(5.872)	(6.256)
Observations	9	9	9	9
R-squared	0.365	0.615	0.365	0.636
Includes quadratic outreach	N	Y	N	Y

*** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses.

Next, we break down the overall non-CIF expenditure into its subcomponents – specifically, CID, STAF, and PM related expenditure. As shown in Figure 4, CID expenditure per member was highest in initial years (at $17.35 per capita), kept declining after 2009, and the decline tampered off after 2012. At the maximum program outreach of 5.7 million members, CID expenditure was between $5.2 (under Model 1) and $6.5 (under Model 2) per capita. The relative share of CID expenditure became closer to PM and STAF related expenditure in the later years of program. STAF related expenditures observed a similar trend, but at much lower levels than the CID expenditure. These costs changed from $3.95 per capita when the project had an outreach of 59,000 members to $0.02 per member when the project had the highest outreach of 5.73 million members. Average expenditure related to PM activities also declined rapidly over time. The relative share of PM related expenses declined to less than 5% of overall project costs by 2012–13 when the program reached over a million households. In the first year of the project, PM expenditure was estimated at $47 per capita, which declined to $0.36 under Model 1 and $0.045 under Model 2 by the latest year (2015-16).

Figure 4. *Jeevika* non-CIF Expenditure Components Over Time



The CID component includes expenditure on project start-up, community mobilization, formation of higher-level federations, capacity management, and implementation of inclusive strategies to ensure that the project activities are prioritized among the most economically and socially deprived communities. Initially, as the program scaled and mobilized more households, a higher proportion of resources was devoted to CID (between years 2008 and 2011). STAF related costs were used to cover expenses related to program innovation, partnerships with government agencies and non-governmental organizations for financial inclusion and technical support on production and livelihoods activities. The project was reported to have spent lower

amounts on the STAF component than initially estimated because of leveraging already established innovation linkages (World Bank, 2018). The inverse U-shaped trend in the relationship between per-member annual CID as well as STAF expenditure and scale suggests increasing expenditure with respect to an increase in the number of households covered when the program is at a relatively small scale (under 200,000 households), and a decrease in costs as the program scales up using its established infrastructure and capacity-building network. Finally, the initial peak in project management expenses may indicate high investment in project implementation support activities in the initial years, including monitoring, evaluation and learning progress and set up of project Management Information System (MIS).

More formally, we find that program scale is negatively associated with all three expenditure components, as shown in regression estimates in Tables 2 through 4. In terms of magnitude, we find that increasing outreach by an additional 100,000 members is associated with the largest decline in PM-related expenditure (by $1.6 to $2.3), followed by CID expenditure ($0.57 to $0.59), while decline in STAF expenditure is slowest with respect to scale (by $0.15 to $0.20). We do see that scale squared is positively associated with average expenditure per member, but this coefficient is not statistically significant under any model, most likely because of a reduction in statistical power. Considering the coefficients from the linear and quadratic outreach variables, the constant estimate indicates that reaching the smallest possible scale costs $14-$15 in CID expenses, $2.2 to $2.4 in STAF expenses, and $19.3 to $20.9 in PM expenses. The same models predict that when the program operates at maximum scale, the per capita expenses reduce to $6.2 to $7.0 for CID, $0.35 to $0.36 for STAF, and $4.4 to $4.8 for PM related components.

Table 2. *Jeevika* non-CIF Expenditure Components: CID

	Model 1 Cumulative Participation		Model 2 Fixed Period Participation	
Reach (in 100,000)	-0.159*	-0.569*	-0.170*	-0.592
	(0.068)	(0.274)	(0.075)	(0.349)
Reach (in 100,000) squared		0.007		0.009
		(0.005)		(0.008)
Constant	12.538***	14.337***	12.979***	14.557***
	(1.725)	(1.964)	(1.587)	(1.993)
Observations	9	9	9	9
R-squared	0.441	0.599	0.422	0.539
Includes quadratic outreach	N	Y	N	Y

*** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses.

Table 3. *Jeevika* non-CIF Expenditure Components: STAF

	Model 1 Cumulative Participation		Model 2 Fixed Period Participation	
Reach (in 100,000)	-0.037*	-0.150*	-0.046*	-0.196*
	(0.019)	(0.075)	(0.022)	(0.097)
Reach (in 100,000) squared		0.002		0.003
		(0.001)		(0.002)
Constant	1.746***	2.242***	1.797***	2.357***
	(0.474)	(0.539)	(0.467)	(0.552)
Observations	9	9	9	9
R-squared	0.367	0.547	0.378	0.561
Includes quadratic outreach	N	Y	N	Y

*** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses.

Table 4. *Jeevika* non-CIF Expenditure Components: PM

	Model 1 Cumulative Participation		Model 2 Fixed Period Participation	
Reach (in 100,000)	-0.313	-1.637	-0.385	-2.305
	(0.247)	(1.042)	(0.299)	(1.328)
Reach (in 100,000) squared		0.024		0.042
		(0.018)		(0.029)
Constant	13.503*	19.309**	13.754*	20.938**
	(6.298)	(7.476)	(6.296)	(7.584)
Observations	9	9	9	9
R-squared	0.187	0.366	0.191	0.407
Includes quadratic outreach	N	Y	N	Y

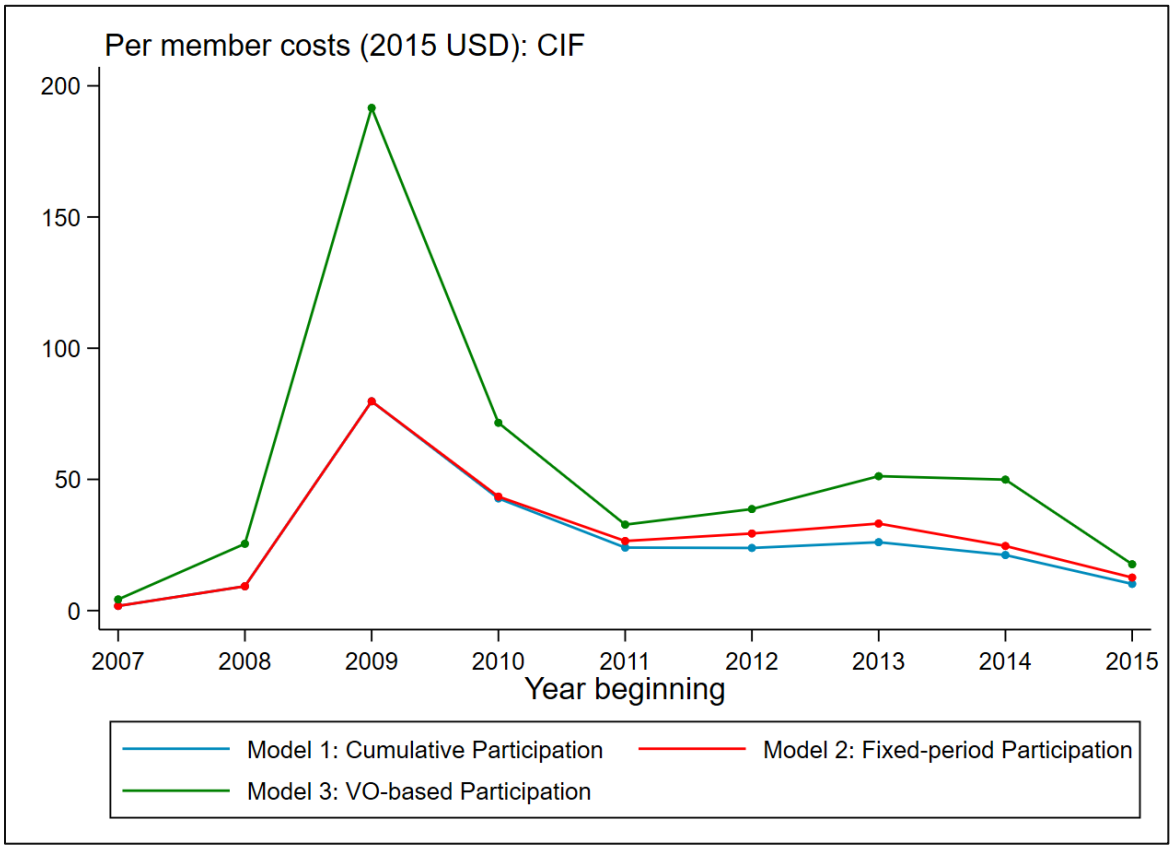
*** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses.

Community Investment Expenditure

Next, we analyze expenditure related to CIF as a separate bucket because of the different activity focus of this component and related assumptions. As shown in Figure 5, expenditure on CIF component was the highest among all project components irrespective of the assumption applied in estimating per member costs. Further, contrary to other expenditure components, CIF expenditure per member did not show a steady decline across all years. Specifically, post *Jeevika*'s expansion after 2011, average CIF expenditure increased over the next four years before declining again between 2014 and 2015 (see Figure 5). While the proportion of CIF expenditure remained highest in all years after 2010, it steadily increased after 2012 (post-NRLM). The trends shown in Figure 5 may be indicative of multiple short-term average cost curves – with economies of scale between 2009 and 2011, followed by diseconomies due to the steep increase in output after 2011. Beyond 2013, we again find evidence for economies of scale as the increase in mobilization slowed down between 2014 and 2015. These funds are used to support financial inclusion (including credit linkage), repayment of high-cost debt, livelihoods initiatives (including practicing new agricultural interventions, formation of producer groups, poultry and non-farm interventions, and formation of micro-enterprises), and interventions related to health, nutrition and sanitation. Group mobilization and linkage to formal banking is the first focus of the program, while CIF-supported activities usually occur among more mature SHGs. Therefore, our estimation of per member CIF expenditure included a third assumption where per member costs are estimated based on members of SHGs that had

federated to form VOs by a given year – as a proxy for SHGs that were eligible to receive CIF support.

Figure 5. *Jeevika* CIF Expenditure Over Time



The trend observed in Figure 5 is consistent with estimates of the regression model summarizing the cost function as shown in Table 5. Model 3, which assumes average cost based on SHG members who were part of VOs, is the only assumption under which scale is negatively associated with per-member costs. The constant term under this model suggests that providing CIF-related activities, at a minimum, cost $65 per member. Overall, we find less evidence for economies of scale under this specific cost component – primarily due to increasing costs related to the steep expansion post 2012. However, we do see (in Figure 5) that the increase flattens out between 2013 and 2014 and further declines beyond 2014, suggesting that per member costs may have continued to fall over time.

Table 5. *Jeevika* CIF Expenditure

	Model 1 Cumulative Participation		Model 2 Fixed Period Participation		Model 3 VO-based Participation	
Reach (in 100,000)	-0.306	0.255	-0.306	1.222	-1.375	-1.360
	(0.404)	(1.916)	(0.494)	(2.479)	(1.834)	(7.003)
Reach (in 100,000) squared		-0.010		-0.034		-0.000
		(0.034)		(0.053)		(0.214)
Constant	31.487**	29.026*	33.076**	27.358	65.413**	65.371*
	(10.298)	(13.745)	(10.388)	(14.154)	(24.610)	(32.581)
Observations	9	9	9	9	9	9
R-squared	0.076	0.089	0.052	0.111	0.074	0.074
Includes quadratic outreach	N	Y	N	Y	N	Y

*** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses

A Discussion on Scale and Implementation: From Costs to Cost-Effectiveness

Our findings suggest that large scale implementation of women's SHG programs has enormous potential to benefit from reduced costs due to economies of scale. The largest source of decline comes from expenditure related to project management and implementation support which include most project monitoring and operational expenses. Other components, specifically mobilization expenditures and expenditures of layering other program activities in form of community investments, also reduce with respect to scale, but the large investments required to support these interventions implies that the negative relationship is observed over a much longer period. It is also noteworthy that CIF expenditure reached a peak two years after project implementation instead of at the beginning of the project like other expenditure components, indicating that the program initially focused on community mobilization and basic financial inclusion activities, and gradually introduced community investments as groups began to form higher level federations.

These findings produce an important side to an often-told incomplete story. Considering the relationship between program impact and scale in isolation from the relationship between program costs and scale masks a crucial element of translating results to policy recommendations and action. To a large extent, *Jeevika* cost analysis differs from other single outcome-focused development interventions because of the scale and complexity of the

program. The program includes multiple components, covers a wide geographical area, and as of 2020, covers over nine million households. The implementation of the program also varies across contexts due to variation in local capacity, population needs, and other contextual factors. Further, as the program scaled up, several components of the program were implemented in collaboration with different organizations. These heterogeneities suggest that the sustainability and momentum of program success may also vary with respect to scale (Mansuri and Rao, 2012). Findings from two impact evaluations (Datta et al. 2015; Hoffman et al., 2020) – the first focusing on the first phase of the project (pre-2012) and the latter on the second phase (2012-2014) indicated that while the project was able to generate strong positive effects, especially on social empowerment, in the first phase, the second phase (post 2012) failed to see these individual-level and household-level effects.

An ethnographic study attributed the failure to produce similar effects in second phase to the pressure to scale up quickly (Majumdar, Rao, & Sanyal, 2017). The study had three primary findings – 1) During the second phase, program participants were mobilized quickly with a fixed script, unlike the first phase which involved creative improvisation that involved multiple community stakeholders; 2) Community facilitators lacked experience and drive in the second phase having spent little to no time on institutional learning; and 3) Limited resources in the second phase meant little room for slow learning, and lower investments in form of capitalization funds.

To test the hypotheses from our model, we examine the extent to which reduced impact is offset by reduced costs. We primarily focus on the costs of achieving benefits related to lower costs of borrowing, estimated as the ratio of costs to household change in high cost debt.[8] In Phase 1, the program led to a 46% decrease in total high cost debt per household, while in phase 2, the impact was approximately 15% (Majumdar, Rao, & Sanyal, 2017; Datta et al., 2015; Hoffman et al., 2020).[9]

As shown in Table 6, we find an interesting pattern in the cost-effectiveness ratios (CERs). Before 2012, that is in phase 1 of the project, the average high cost loan amount decreased by

[8] *Jeevika*'s impact on reduced costs of borrowing are expected to increase with scale if more individuals are able to access low-cost credit, both as a result of direct program impact and as a result of general equilibrium effects leading to lower interest rates charged by informal moneylenders. The effects reported by Hoffman et al. (2018) include the change in interest rate from both channels.

[9] Because the two evaluations were conducted over four years apart, the definition of "high cost" varied across the two studies. Phase 1 study defined high cost loans as loans with a monthly interest rate higher than 2%, while Phase 2 study considered loans with monthly interest rate higher than 4% as high cost loans.

$85 because of the program, more than twice the average impact of $33 in phase 2. However, it costed 66 cents to reduce each additional dollar of high cost debt in phase 1 – only slightly different from 73 cents in phase 2 when considering expenditure on CID and STAF only (row A). This estimation does not include Project Management expenditures, because the bulk of this expenditure component included monitoring and implementation support, which was disproportionately spread in the beginning of the project. When we do add PM expenses (row B), findings indicate that it costed $0.92 to produce an additional unit of impact in phase 1 – not meaningfully different from $0.89 in phase 2. Finally, when we add Community Investment Funds (row C), we find a cost-effectiveness ratio of 2.5 in phase 1, lower than 3.2 in phase 2.

Overall, the analysis suggests that stark differences between impacts on high cost loans of households during the pilot phase and the scaled-up program fade away when we estimate cost-effectiveness ratios. Household benefits resulting from a decrease in high cost loans when the project was implemented at a small scale (phase 1) were more than twice the size of the benefits during the scaled-up implementation (phase 2). The program shows a much smaller difference in cost-effectiveness ratios for reduced high cost debt when comparing the scaled-up phase with the pilot phase of the program.

Table 6. Cost Per Dollar Change in High Cost Loans

		Phase 1	Phase 2
	Average change in high cost loan amount per household (2018 USD)	*85*	*33*
	Average expenditure for duration of evaluation (2018 USD)		
	CID costs per capita	50	22
	CIF costs per capita	133	76
	STAF costs per capita	6	2
	PM costs per capita	22	5
(A)	Cost per dollar of benefit (CID and STAF costs only)	0.66	0.73
(B)	Cost per dollar of benefit (all non-CIF costs)	0.92	0.89
(C)	Cost per dollar of benefit (all costs)	2.49	3.21

Note: All costs and benefits were converted to 2018 USD by first adjusting for Rupee inflation using the CPI method, and then converting to 2018 US Dollars using the Market Exchange Rate. Assumptions include a uniform average loan amount of Rs.10,000 in both phases for both treated and control groups, estimated from Hoffman et al. and Datta et al.

Some caveats should be considered when interpreting these findings. First, the Phase 1 evaluation is less robust as it relied on propensity score matching on retrospective data, instead of a randomized experiment, while the phase 2 evaluation used a randomized controlled trial design. Therefore, estimates of impact from Phase 1 evaluation may be an overestimate of the true program effect if otherwise motivated women were more likely to participate in the program.[10] Second, the cost-effectiveness ratios shown here only apply to benefits caused by reduced costs of borrowing resulting from reduced interest rates. For this study, our primary aim is to demonstrate the consequences of ignoring cost data when interpreting findings of impact evaluations of scaled-up programs. Several other outcomes need to be considered in the estimation of a holistic benefit-to-cost ratio.[11] The Phase 1 evaluation showed that the program led to significant improvements in women's empowerment outcomes including mobility, decision-making, and propensities toward collective action; while the Phase 2 evaluation showed only weak and mostly conflicting effects on empowerment indicators (except for landless households).

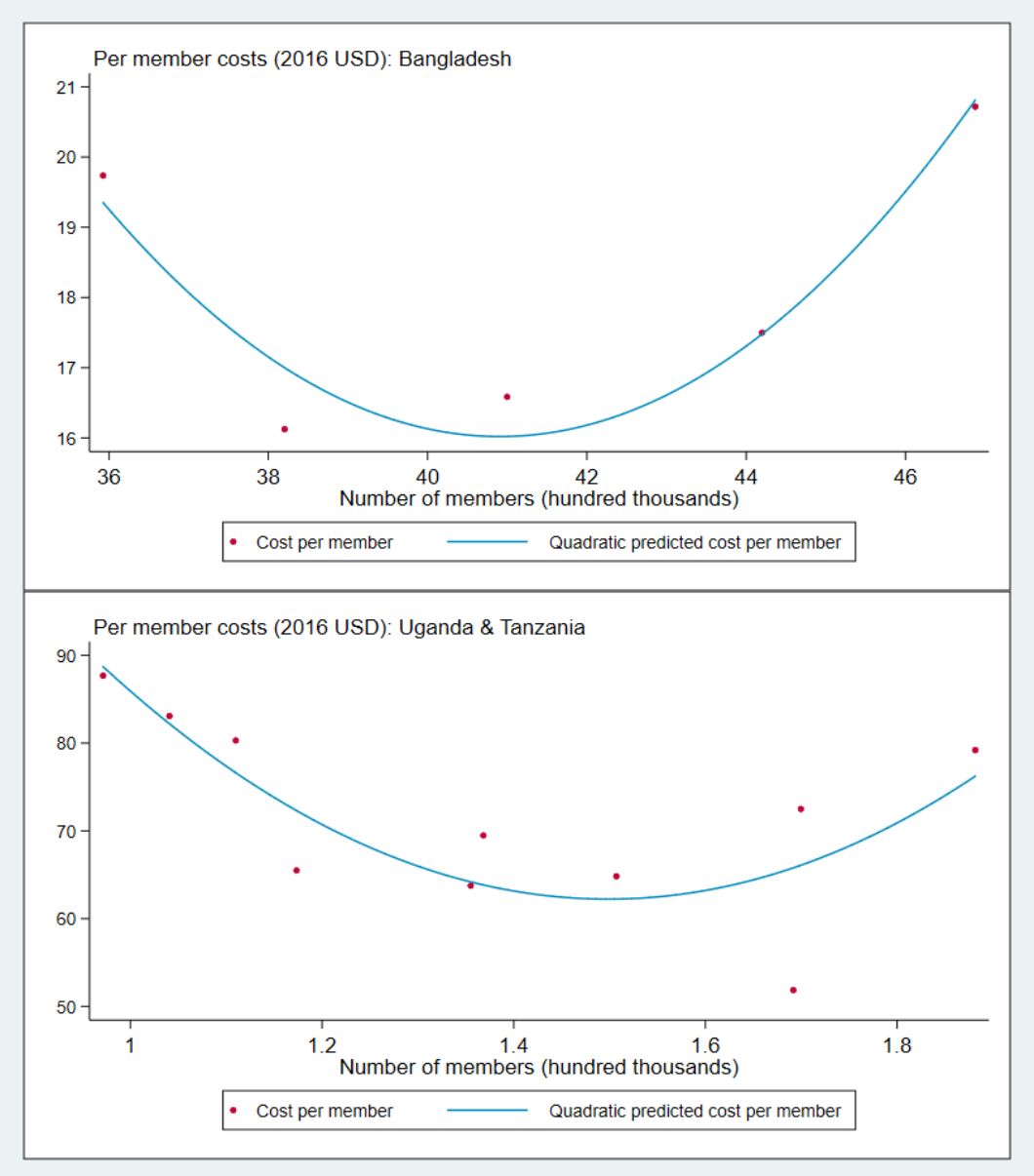
Robustness Checks on Cost-Efficiencies

As a secondary analysis and to substantiate our findings on cost-efficiencies as a function of program scale, and to assess the external validity of the results, we used annual data on program expenditure and outreach of BRAC's savings group-based microfinance program in Bangladesh, Uganda, and Tanzania. We used these data as a robustness check because of data availability but also because BRAC's microfinance programs are amongst the largest women's group programs with economic objectives in the world and operate across several countries. BRAC (originally Bangladesh Rehabilitation Assistance Committee) is among the world's largest non-profit organizations with a scale of over 120 million people across eleven countries (BRAC, 2017). BRAC's groups-based microfinance program, which offers group-based loans delivered exclusively to women, started in Bangladesh in the 1970s, and by 2017, had reached over eight countries. We collected data on operating costs per loan disbursement between 2013 and 2017 in Bangladesh (where the program had been operating for a long time and at a large scale), Uganda, and Tanzania. Figure 6 shows the operational costs per borrower with respect to total number of borrowers in the three countries. Between 2013 and 2017,

[10] Brody et al. (2015) show that quasi-experimental evaluations of SHGs with a high risk of bias tend to overestimate the impact of SHGs on women's empowerment in a meta-analysis.
[11] We will present these benefit-cost ratios in a forthcoming study that includes longer-term impact estimates of the Jeevika program (Paul et al., *in progress*).

average cost per borrower was $18.13 in Bangladesh, $70.67 in Uganda, and $72.97 in Tanzania. At the same time, BRAC microfinance had a much larger scale in Bangladesh, which is not surprising given that BRAC has operated in Bangladesh for almost four decades, while the operations in Tanzania and Uganda started much more recently in 2006.[12]

Figure 6. BRAC Microfinance Costs: 2013 to 2017



[12] Specifically, in Uganda, there was a spike in operational costs between 2015 and 2016. BRAC's annual reports suggest that this increase is likely due to additional expenses for the transformation of BRAC microfinance to a regulated credit institution (tier II) company (BRAC, 2016; 2017).

To analyze returns to scale, we examined how cost per borrower changed with the number of borrowers in each year (a proxy for microfinance outreach). Table 7 shows the formal regression estimates. The bivariate relationship between scale and costs (Column 1) indicates that an increase in member outreach of 100,000 persons is associated with a $1.3 decline in operational costs per borrower (p<0.01). Column 2 adds the quadratic outreach term and shows that the decline in costs with respect to scale now increases to $3.1 for every 100,000 additional persons, but the decline slows down by $0.04 (although insignificant). In column 3, we add country fixed effects, and find that after controlling for country-level differences, an additional 100,000 persons is associated with a decline in per member cost of $20, and that the decline slows down by $0.25 per 100,000 increase in outreach. The regressions show a qualitatively similar pattern as in the costing analyses of the *Jeevika* program, suggesting that the results may well be externally valid.

Table 7. BRAC Microfinance: 2013 to 2017

	(1) Cost per borrower (2015 USD)	(2) Cost per borrower (2015 USD)	(3) Cost per borrower (2015 USD)
Number of borrowers (in 100,000)	-1.328***	-3.147**	-20.345*
	(0.132)	(1.133)	(11.119)
Number of borrowers (in 100,000) squared		0.042	0.247*
		(0.026)	(0.135)
Country FEs: Tanzania			-335.777
			(216.228)
Country FEs: Uganda			-331.387
			(212.589)
Constant	73.400***	76.144***	432.986*
	(3.152)	(3.424)	(229.380)
Observations	15	15	15
R-squared	0.887	0.907	0.925
Includes quadratic outreach	N	Y	Y
Includes country FEs	N	N	Y

*** p<0.01, ** p<0.05, * p<0.1 Standard errors in parentheses

Conclusion

This study brings together two seemingly disjoint streams of discourse in translating evidence from impact evaluations to policymakers – First, the past decade has seen an increasing

emphasis, especially in international development, on incorporating cost estimates and cost-benefit analyses into impact evaluations (The World Bank, 2010; Brown & Tanner, 2019). Second, while policy and program evaluations of development interventions often show positive impacts of pilot programs, researchers have demonstrated that pilot programs may fail to achieve similar success when scaled up (Vivalt, 2020; Bold et al., 2018). Yet, there is currently little to no discussion on how large-scale interventions may also benefit from economies of scale in terms of lower implementation costs – a factor that is especially crucial for group-based programs like SHGs which are able to deliver services to many people at once (Diaz-Martin et al., 2020).

Focusing on the *Jeevika* SHG program in Bihar, India, we show that large-scale development interventions can achieve significant economies of scale from expanding program outreach. Specifically, we found that per capita expenditure on basic program activities (all non-CIF components) declined from $37 in the program's first year when it covered approximately 6,000 members to $13 after eight years when it reached approximately 5.7 million members. These costs declined by $3 for every 100,000 new households, but the decline slowed down with a large increase in scale. We separately analyzed expenditure on the CIF component that included community grants for livelihoods activities and found a similar negative association between scale and costs, although the estimates were not statistically significant. CIF related expenditure differs from expenditure on other program activities because the receipt of different fund components under CIF often depends on SHG maturity, whether they had a microenterprise plan etc. Upon combining these cost estimates with program impact from the pilot phase and the scaled-up phase, we found that average high cost loan amount decreased by $85 in phase 1, more than twice the average impact of $33 in phase 2. However, this difference in program impact did not translate to differences in cost-effectiveness, because the lower implementation costs of the scaled-up model were able to mostly offset the lower benefits. Additional analyses of the operational costs of BRAC's group-based microfinance program in Bangladesh, Uganda, and Tanzania show a similar significant decline with respect to program scale, suggesting that the findings on economies of scale are likely externally valid.

Studies have shown that implementing development interventions at scale is challenging. In a recent note on insights from research in development economies, Artuc et al. (2019) indicated that as programs grow, they are often met with political economy issues arising from low state capacity or poor bureaucratic management. In addition, it is likely that programs that generate positive effects are highly resource-intensive, and such resources are simply not available to

operate the program at scale. Indeed, the expansion of *Jeevika*'s SHG program faced similar challenges. Along with the implementation challenges noted earlier, *Jeevika* also decreased the value of the Initial Capitalization Fund (ICF) from INR 50,000 to INR 15,000. The ICF is one component of the CIF funds, and is provided to SHGs after they attain maturity in four to six months. SHG members can take turns to borrow out of the ICF pot for a variety of reasons, such as consumption smoothing, debt reduction and productive investments. Reduction in ICF amounts may imply that groups were unable to take on a similar range of activities in Phase 2 as in Phase 1.

While scaled-up programs may show smaller effects, they may also benefit from economies of scale, indicating that an increased use of cost-effectiveness analyses in impact evaluations is only the first step when making decisions about the scale-up of pilot programs in international development. It is not only critical to examine the costs and cost-effectiveness of development programs, but also to examine how costs and cost-effectiveness change with the scale-up of development programs. Such analyses require longitudinal data on the reach, costs, and impact of development programs, and how these change with the number of program participants. Such data are currently very scarce, which makes it challenging for decision makers to make evidence-based decisions about the scale-up of international development programs. The relationship between scale, costs, and cost-effectiveness will also depend strongly on the intervention type and contextual characteristics. While the *Jeevika* program in Bihar may show a similar cost-effectiveness ratio after scale-up as in its pilot phase, the relationship between costs, cost-effectiveness, and scale may be very different for other women's group programs with economic objectives or more generally for international development programs.

Regardless, however, our results show that the current emphasis on the reduced effectiveness after program scale-up is not sufficient and may result in a too pessimistic view about the possibility of achieving positive impacts at scale. While program effectiveness may reduce after scale-up for most international development programs (Vivalt, 2020), the presence of economies of scale may justify the scaling of successful pilot programs even when their impact reduces with the number of program participants.

Clearly, more research is needed to examine how program costs and cost-effectiveness of international development programs change with scale. Such research may be able to address some of the limitations of our study as well. For example, the finding that *Jeevika* led to a reduction in interest rates charged by informal moneylenders (Hoffman et al., 2018; Kochar et

al., 2020), may imply that as the program scales up further, the extent of these general equilibrium effects will be realized over a much wider area, beyond the program villages. Given the large scale of NRLM's operations across the country, it will be critical to examine these general equilibrium effects and their relationship to the scale-up of the program in more detail. Such studies should into take into consideration how such general equilibrium effects influence inequality in economic outcomes considering that scaled-up programs reach a much larger number of women than pilot programs. Further, future studies should focus on collecting more disaggregated data to estimate costs related to specific activities like farm and non-farm-based livelihoods as well costs of resources which were not directly paid for. More rigorous implementation research on *Jeevika* would also enable for the estimation of cost-effectiveness ratios based on more detailed information about the number of SHGs that were eligible to receive *Jeevika*'s support on different activities in every year. Finally, future research should focus on estimating the program's cost-effectiveness with respect to multiple outcome domains to better compare costs against outcomes that policymakers may hold highest value for.

References

Artuc, E., Cull, R., Dasgupta, S., Fattal, R., Filmer, D., Giné, X., ... & Kraay, A. (2020). Toward Successful Development Policies: Insights from Research in Development Economics.

Bold, T., Kimenyi, M., Mwabu, G., & Sandefur, J. (2018). Experimental evidence on scaling up education reforms in Kenya. *Journal of Public Economics*, *168*, 1-20.

BRAC. (n.d.). *BRAC Annual Reports 2013-2017.* Retrieved online from http://www.brac.net/annual-reports-and-publications

Breza, E., & Kinnan, C. (2018). *Measuring the equilibrium impacts of credit: Evidence from the Indian microfinance crisis* (No. w24329). National Bureau of Economic Research.

Brody, C., De Hoop, T., Vojtkova, M., Warnock, R., Dunbar, M., Murthy, P., & Dworkin, S. L. (2015). Economic Self-Help group programs for improving women's Empowerment: a systematic review. *Campbell Systematic Reviews*, *11*(1), 1-182.

Brown, E. D., & Tanner, J. C. (2019). *Integrating Value for Money and Impact Evaluations: Issues, Institutions, and Opportunities*. The World Bank.

Bryan, G., Chowdhury, S., & Mobarak, A. M. (2014). Underinvestment in a profitable technology: The case of seasonal migration in Bangladesh. *Econometrica*, *82*(5), 1671-1748.

Colbourn, T., Pulkki-Brännström, A. M., Nambiar, B., Kim, S., Bondo, A., Banda, L., ... & Costello, A. (2015). Cost-effectiveness and affordability of community mobilisation through women's groups and quality improvement in health facilities (MaiKhanda trial) in Malawi. *Cost Effectiveness and Resource Allocation*, *13*(1), 1.

Datta, U. (2015). Socio-economic impacts of JEEViKA: A large-scale self-help group project in Bihar, India. *World Development*, *68*, 1-18.

Datta, Upamanyu; Rao, Vijayendra. (2018). Jeevika Livelihoods Project, Phase 1 Evaluation. One round "retrospective" evaluation. Household Survey Data 2011, https://doi.org/10.7910/DVN/QHF29P. Harvard Dataverse, V1.

Datta, Upamanyu; Rao, Vijayendra. (2018). Jeevika Livelihoods Project Phase 2 Evaluation (RCT), Bihar, India - Baseline and Endline Household And Village Data 2011-2014. https://doi.org/10.7910/DVN/6PAHVM. Harvard Dataverse, V1

Deininger, K., & Liu, Y. (2013). Economic and social impacts of an innovative self-help group model in India. *World Development, 43*, 149–163.

Dìaz-Martin, L., Gopalan, A., Guarnieri, E., & Jayachandran, S. (2020). Greater than the Sum of the Parts? Evidence on Mechanisms Operating in Women's Groups. Working Paper. https://faculty.wcas.northwestern.edu/~sjv340/womens_groups.pdf

Government of India. (2011). Census 2011 results. Retrieved September 2019 from http://censusindia.gov.in/2011-prov-results/data_files/india/Final_PPT_2011chapter7.pdf?_sm_au_=iVVS1tPffSqnVskj

Harper, M. (2002). Promotion of self help groups under the SHG bank linkage programme in India. Presented at the Seminar on SHG-Bank Linkage Programme, New Delhi, India.

Hoffmann, V., Rao, V., Surendra, V., & Datta, U. (2020). Relief from usury: Impact of a self-help group lending program in rural India. Journal of Development Economics, 102567.

International Institute for Population Sciences (IIPS) and ICF. (2017). National Family Health Survey (NFHS-4), 2015-16: India. Mumbai: IIPS. Retrieved September 2019 from https://dhsprogram.com/pubs/pdf/FR339/FR339.pdf

Isern, J., Prakash, L. B., Pillai, A., Hashemi, S., Christen, R. P., Ivatury, G., . . . Rosenberg, R. (2007). Sustainability of self-help groups in India: Two analyses (Occasional Paper No. 12). Washington, DC: Consultative Group to Assist the Poor (CGAP).

Jeevika. (n.d.) *Jeevika Annual Reports 2006-07 to 2015-16*. Retrieved February 2019 from http://brlp.in/web/brlp/annual-audit-reports

Johns, B., & Torres, T. T. (2005). Costs of scaling up health interventions: a systematic review. *Health policy and planning*, *20*(1), 1-13.

Kochar, A., Barooah, B., Jain, C., Closepet, N., Narayanan, R., Sarkar, R., & Shah, R. (2020). *Impact Evaluation of the National Rural Livelihoods Programme*. Retrieved from https://www.3ieimpact.org/sites/default/files/2020-10/IE128-NRLP-NRLM.pdf

Majumdar, S., Rao, V., & Sanyal, P. (2017). *On the frontlines of scaling-up: a qualitative analysis of implementation challenges in a CDD project in rural India*. The World Bank.

Mangham-Jefferies, L., Pitt, C., Cousens, S., Mills, A., & Schellenberg, J. (2014). Cost-effectiveness of strategies to improve the utilization and provision of maternal and newborn health care in low-income and lower-middle-income countries: a systematic review. *BMC pregnancy and childbirth*, *14*(1), 243.

Mansuri, G., & Rao, V. (2012). Can participation be induced. *Some evidence from developing countries, A World Bank Policy Research Working Paper*, (6239).

Ministry of Rural Development. (2011). *Natural Rural Livelihoods Mission programme implementation plan.* New Delhi, India: Government of India.

Muralidharan, K., Niehaus, P., & Sukhtankar, S. (2020). *Identity verification standards in welfare programs: experimental evidence from India* (No. w26744). National Bureau of Economic Research.

Paul, S., Siwach, G., de Hoop, T, Jain, C., & Barooah, B. (Forthcoming). Costs and Cost-Effectiveness of the Jeevika Program in Bihar.

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

Pulkki- Brännström, A-M., Haghparast-Bidgoli, H., Batura, N., Colbourn, T., Azad, K., Banda, F., Banda, L., Borghi, J., Fottrell, E., Kim, S., Makwenda, C., Kumar Ohja, A., Prost, A., Rosato, M., Kumer Shaha, S., Sinha, R., Costello, A., & Skordis, J. (2020). Participatory learning and action cycles with women's groups to prevent neonatal death in low-resource settings: A multi-country comparison of cost-effectiveness and affordability. *Health Policy and Planning.*

Shylendra, H. (2018). Federating the Self-Help Groups (SHGs) in India: Emerging Conceptual and Policy Challenges. *International Journal of Rural Management 14*:154-181. Retrieved September 2019 from https://journals.sagepub.com/doi/abs/10.1177/0973005218798199

Sinha, R. K., Haghparast-Bidgoli, H., Tripathy, P. K., Nair, N., Gope, R., Rath, S., & Prost, A. (2017). Economic evaluation of participatory learning and action with women's groups facilitated by Accredited Social Health Activists to improve birth outcomes in rural eastern India. *Cost Effectiveness and Resource Allocation*, *15*(1), 2.

Tankha, A. (2002). Self-help groups as financial intermediaries in India: Cost of promotion, sustainability and impact. The Hague, The Netherlands: Cordaid.

Tulloch, C. (2019). Taking intervention costs seriously: a new, old toolbox for inference about costs. *Journal of Development Effectiveness*, *11*(3), 273-287.

Venton, C. C., Tsegay, E., Etherington, K., Dejenu, M., & Dadi, T. (2013). Partnerships for change: A cost benefit analysis of self help groups in Ethiopia. Addis Ababa, Ethiopia: Tearfund.

Vivalt, E. (2020). How Much Can We Generalize from Impact Evaluations? *Journal of the European Economic Association.*

World Bank. (2007). *Project Appraisal Document for a Bihar Rural Livelihoods Project – Jeevika.* Retrieved February, 2019 from http://documents.worldbank.org/curated/en/675101468269084791/pdf/39039.pdf

World Bank. (2012b). *Implementation Status & Results: Bihar Rural Livelihoods Project – Jeevika.* Retrieved March, 2019 from

http://documents.worldbank.org/curated/en/953391468771088485/pdf/ISR0Disclosabl0202020120140169735205.pdf

World Bank. (2017). *A Decade of Rural Transformation: Lessons Learnt from the Bihar Rural Livelihoods Project – Jeevika.* Retrieved February, 2019 from http://documents.worldbank.org/curated/en/298391515516507115/pdf/122548-WP-P090764-PUBLIC-India-BRLP-Booklet-p.pdf

World Bank (2018a). *International Development Association project appraisal document on a proposed credit to the Federal Republic of Nigeria in the amount of 69.6 SDR million (US$100 million equivalent) for a Nigeria for Women Project.* Retrieved from http://documents1.worldbank.org/curated/en/207671530329469779/pdf/NIGERIA-PAD-05252018.pdf

World Bank. (2018b). *India State Briefs: Bihar.* Retrieved September 2019 from https://www.worldbank.org/en/news/feature/2016/05/26/india-states-briefs

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