ECWG
EVIDENCE CONSORTIUM ON WOMEN'S GROUPS

ECWG WORKING PAPER 02 UPDATED IN JUNE 2021

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

Contents

Abstract

Financial inclusion and livelihoods programs, like microfinance, financial literacy, and women's empowerment programs, often show reduced impacts when scaled up. Yet, scaling up may lead to economies of scale, which can reduce average costs of operations and maintain the program's 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 program in India changed over a twelve-year period. The study focuses on program costs of *Jeevika* – the Bihar Rural Livelihoods Promotion Society – from its inception in 2007 to 2019. Using expenditure data from program's audit statements, we estimate annual program costs of different components, including capacity building, institutional development, and project management. Our study finds that annual per capita expenditure on basic program activities declined from $34 per program participant when the program covered approximately 60,000 members to $3 per program participant when it reached over 11 million members. Incorporating existing information on program impact from the pilot and scaled-up phases, we found that the program led to reductions in household debt from high-cost loans by $89 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. Additionally, we found that formation of higher-level federations can be achieved at a relatively lower marginal cost. Given the low costs and their positive role in supporting community institutions for women, we suggest that groups should be linked to federations to further improve overall cost-effectiveness. Finally, we argue that program implementers must identify key success factors in pilot stages of programming to minimize tradeoffs between cost savings and potentially reduced impacts at scale.

Keywords: Costing; Scaling up; Cost-effectiveness; Self-Help Groups; Federations; India

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 were traditionally 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) has mobilized over 70 million households into SHGs with the goal of providing them with access to financial services and sustainable livelihoods enhancements (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 group-based savings, credit, and livelihoods programs in six states of the country (World Bank, 2018a).

To some extent, these decisions to scale are justified by evidence on the impact of women's groups. A systematic review showed that SHGs 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). Apart from economic SHGs, research also shows that open, community groups of women that involve participatory learning 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). 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. While a few studies from different contexts do present program costs when evaluating the program impacts (Deininger & Liu, 2015; Karlan, et al., 2017), the contextual determinants of costs remain unclear. Given the growing interest in expanding the outreach of these programs, it is imperative to study the potential role of scale in influencing program impact, implementation, and costs.

Program scale may impact outcomes through different mechanisms. On the one hand, increases in scale may lead to an adverse impact on implementation quality leading to lower effectiveness. A recent meta-analysis of development programs, including those that focus on microfinance, financial literacy and women's empowerment, concluded that larger programs and programs that were implemented by governments had lower effects (Vivalt, 2020). Indeed, program-specific evidence on other development interventions suggests 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 et al., 2014). More recent evidence indicates that prior experience in communities could explain the larger positive effects of interventions implemented by NGOs relative to interventions implemented by governments (Jeuland et al., 2020).

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). Previous evidence from health programs and interventions aiming to reduce gender-based violence indeed show that the costs of these programs reduce significantly when they reach a larger number of program participants (Abramsky et al., 2012; Bijlmakers et al., 2018; Brooker et al., 2008; Jan et al., 2011). 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 this 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 11 million women into 0.9 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. Using program expenditure data from annual audit reports of *Jeevika* between 2007-08 and 2018-19, we estimate costs of different program components, including capacity building, institutional development, and project management. We use these estimated costs to empirically test for evidence of economies of scale. In addition, we provide evidence on how costs of the *Jeevika* program change with the inclusion of extra programming, and how program cost-effectiveness may change with 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 a good 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 & Torres, 2005). While a large dataset with data on costs of multiple programs over many periods and program components would be ideal, such data is almost never readily available. Instead, most studies observe cost data at one point in time or have limited data on the costs of different program components precluding researchers from studying how costs vary with changes in both program scale and program scope. Therefore, even a small increase in the number of cost data points over time can significantly increase the evidence-base on how costs change with scale, especially when data include costs of different program components. 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 SHGs deliver an increased number of activities. For example, evidence indicates 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 though some evidence indicates that group-based models can indeed achieve the same outcomes at a lower cost than programs targeting individuals. For example, Grantham-McGregor et al. (2020) demonstrate that home visiting and group sessions had similar impacts on early childhood cognition in Odisha, India, but group sessions were 3.5 times less expensive than home visits.

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., 2018; Majumdar et al., 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., 2018). Other studies suggest that scaled-up programs may have larger general equilibrium effects through changes in prices and wages (Breza & Kinnan, 2018; Muralidharan et al., 2020). The evaluation of the scaled-up *Jeevika* program suggests that the program led to a reduction in interest rates charged by informal money lenders, likely due to increased access to formal credit from *Jeevika* and reduced demand for informal credit (Hoffman et al., 2021; Kochar et al., 2020). The reductions in interest rates resulted in particularly large benefits for marginalized populations, such as scheduled caste (SC) and scheduled tribe (ST) households and landless households. The program reduced the difference in average borrowing rates for landholding and landless households by two thirds, and had positive effects on ownership of productive and consumption assets for landless households (Hoffman et al., 2018).

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. While the dataset remains small, the time series cost data enable us to distinguish between short-run and long-term costs during the pilot and after the scale-up of the program.

The results show evidence for substantial economies of scale that increase with the formation of federations at higher levels. Our findings show that the annual per-household expenditure for basic program activities under *Jeevika* declined from approximately $34 per member when the program served fewer than 60,000 households, to almost $3 per member at its scaled-up level when the program reached more than 11 million households. Additionally, the results indicate larger economies of scale for the formation of higher level SHG federations, which is central to the implementation of the *Jeevika* program. SHGs under *Jeevika* follow a federated structure with functional integration between groups at three levels (SHGs, Village Organizations, and Cluster-Level Federations), with SHG as the basic unit. We find that a 1 percent increase in the number of SHG federations formed is associated with an increase in program expenditures of only 0.49 percent.[1]

We also find that the scaled-up version of *Jeevika* may be as cost-effective in achieving benefits from reduced dependence on high-cost loans as its pilot version because of economies of scale. We examined this question by assessing how the cost-effectiveness of *Jeevika* could change with scale by comparing a cost-effectiveness ratio (CER) for benefits from reduced dependence on high-cost loans after an initial pilot of *Jeevika* with an estimate of the CER after the scale up of the program. While the scaled-up program may be equally cost-effective as the pilot in achieving benefits from reduced dependence on high-cost loans, the scaled-up program is, on average, less cost-effective than the pilot in achieving positive effects on women's empowerment and asset ownership. This paper treats reduced dependency on formal loans as the primary outcome measure, but we also examined potential downstream effects on women's empowerment and ownership of consumption and productive assets. Unlike during the pilot, the evaluation of the scaled-up *Jeevika* program did not find statistically significant average effects on women's asset ownership and decision-making power outcomes (Datta, 2015; Hoffman et al., 2018). Yet, the evaluation of the scaled-up program showed evidence for heterogeneous

[1] 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.

effects, with landless households benefiting more from lower costs of borrowing, which led to increased asset ownership for this sub-group. Specifically for landless households, we found that the scaled-up program costed $160 for a one standard deviation increase in the consumption asset index, and $722 for a one standard deviation increase in the productive asset index.[2]

Conceptual Framework

Considering a classical production function framework, *Jeevika* operates with economies of scale when average costs reduce with the number of women mobilized, and diseconomies of scale when average costs increase with the number of women mobilized.[3] While standard microeconomic theory suggests lower average costs after an increase in scale, the magnitude of economies of scale for a program like *Jeevika* also depends on the complexities of program delivery tied to the political economy of the implementation of government programs, including but not limited to government bureaucracy (Artuc et al., 2020), as well as the group-based nature of the program (Diaz-Martin et al., 2020).

Political economy considerations indicate that complex multi-faceted programs, such as *Jeevika*, require effective interagency coordination to achieve complementarities with other government programs and departments (Slater et al., 2016). A high number of policy actors may result in conflicting incentives (Lipsky, 2010; Pressman & Wildavsky, 1980), and implementation over an extended period may result in changing program priorities over time (Bruns et al., 2019). Complex multi-faceted programs also require more contextual information for effective implementation (Ricks & Doner, 2021), which could result in increased costs of targeting and adapting the programming to newly added geographies, especially with increased diversity at the district level. Additionally, scaling up to new districts is usually accompanied by changes in institutional capacity, which may lead to differences in implementation and effectiveness (Paul et al., 2021). Finally, institutional theorists have posited that “visible” programs may receive broader support if politicians receive credit for successful implementation (Paul et al., 2021). One hypothesis is that public programs have incentives to overspend, because expenditures are a direct visible outcome that could proxy for the program’s success in the absence of impact

[2] We could, unfortunately, not estimate the same effects for the pilot phase because the pilot study does not report heterogenous effects for landless households.
[3] We recognize that even in the presence of increasing returns to scale, marginal costs may increase while average costs continue decreasing, but this paper focuses on average costs to account for economies of scale in line with Tulloch et al. (2019).

data. Nonetheless, almost all government programs continue to have incentives to maximize concrete project outcomes within a budget constraint because of limited resources.

Looking at informal micro-level institutional factors, both costs and impacts depend on existing community norms and collective preferences of the household. Understanding local context is important to understand the external validity of the results (Deaton & Cartwright, 2018). On the impact side, Garikipati (2008) argues that a non-patriarchal hold on productive assets is a necessary condition for positive impacts of SHGs on women's empowerment in India. An evaluation of the National Rural Livelihoods Project in seven states including Bihar further found that men's education affects loan amounts drawn by SHGs, and their husband's occupation explains women's decision-making and labor force participation outside the household (Kochar et al., 2020). SHG programs also may have higher benefits for *a priori* more marginalized populations – for example, scheduled tribes in Andhra Pradesh (Prennushi & Gupta, 2014), and landless households (predominantly scheduled caste) in Bihar (Hoffmann et al. 2018).

A recent study on costs of economic inclusion programs found comparable average program costs of *Jeevika* in Bihar, and the NRLM which is implemented countrywide (Paul et al., 2021). Yet, we expect significant heterogeneities within states because of varying social norms that may lead to higher targeting and program adaptation costs. Participation in SHGs, as well as some of the intended outcomes, such as women's mobility, often require the transgression of gender and social norms (De Hoop et al., 2014). Conducting outreach to maximize program participation, therefore, may require additional time and costs. Qualitative studies have noted time investment as a crucial factor for effective implementation of *Jeevika*. In the initial stages of program mobilization, coordinators spent time collecting information on underlying community practices including caste dynamics, land use patterns, informal moneylending, and patriarchal norms, and changed their approaches for mobilization when they realized conflicts with existing norms and practices (Hoffman et al., 2018). For example, coordinators spent extensive time on negotiations and getting buy-in from male members and other members of the household and village, understanding the culture of caste-based handouts to differentiate *Jeevika*'s rollout form existing practices, and addressing community-specific concerns (Hoffman et al., 2018).[4] These processes of effective engagement would likely increase program costs when expanding to new diverse areas.

[4] For example, Hoffman et al. (2018) reported that in some villages, program coordinators received pushback by the Muslim religious head, on the practice of accruing interests on loans. The coordinators adapted their targeting messages, considering these community-specific norms and associated challenges.

One other key institutional feature that may affect the magnitude of economies of scale of *Jeevika* is the group-based nature of the program. As indicated by Diaz-Martin et al. (2020), SHGs and other women's groups may achieve economies of scale because it is less costly to deliver training to the same number of women in a group setting than through individual home visits. In addition, collective action may generate larger impacts when groups enable members to gain access to new markets and services or members can achieve larger collective bargaining power (Anderson et al., 2019).

In this study, we start from the premise that *Jeevika* aims to maximize women's social and economic empowerment at the lowest cost; or in other words, the program aims to minimize operational costs for each level of outcome. 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 program exposure and a 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. An ethnography study found substantial differences in the quality of implementation of *Jeevika* before 2012 when the program operated at a smaller scale, and after 2012 when the program scaled up across the state (Majumdar et al., 2017).

While economies of scale suggest that costs may be lower when program expands beyond its initial outreach, the institutional factors listed above indicate that the scope of cost efficiencies depends heavily on institutional mechanisms. In subsequent sections, we study how average costs of the program changed with respect to scale and shed some light on the question of implications for program cost-effectiveness.

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, 2018b). The state 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, 2018b). Data from the National Family Health Survey-4 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 accounts 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,[5] 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. 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, and by 2021, had mobilized over eleven million households.[6]

The federated structure of SHGs under *Jeevika* facilitates collective action, adoption of livelihoods enhancement and income generating activities, and development of linkages with market institutions (World Bank, 2017). A typical *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., 2018). In the second tier, SHGs are federated into Village Organizations (VOs), which are further federated into Cluster Level Federations (CLFs) at the third level.

Program expenditures under *Jeevika* are reported for four broad components: Community Institution Development (CID), Special Technical Assistance Fund (STAF), Project Management (PM), and the Community Investment Fund (CIF). The CID fund provides all expenses for

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

[6] 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. 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).

institution building, including the costs of mobilizing groups, developing direct linkages to the formal finance services providers, formation of federations, and capacity building of staff at the district, block, and cluster level. The STAF provides technical assistance to the formal financial sector and other partners 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. Overall coordination, project implementation, financial management and monitoring and evaluation at the state and district levels are supported through resources under the Project Management (PM) component. Finally, 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. The VRF is usually provided to VOs to address risks, including food security risks, health risks, and natural disasters through the Food Security Fund or Health Security Fund. *Jeevika* disburses CIF to SHGs and federations as a grant on a demand-driven basis through a participatory micro-planning process.

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 2018-19. 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. It is important to note that our analyses are only based on reported program expenditures. Ideally, a costing analysis should 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.[7,8]

***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. We accessed and analyzed data from two impact evaluations of the *Jeevika* program – the first focusing on the pilot phase of the project pre-2012 (Datta, 2015) and the latter on the second scaled-up phase after 2012 (Hoffman et al., 2018; 2021). Where needed, we complemented the estimates from the published literature with data posted on Harvard Dataverse (Datta & Rao, 2018a; 2018b). The phase 1 survey dataset included 3,997 observations, 50% of whom belonged to treatment areas where *Jeevika* entered in 2008. The phase 2 dataset included 5,870 observations, 50% of whom belonged to treatment areas where *Jeevika* entered in 2012.

Methods

We examine the costs of increasing both the scope as well as the scale of services related 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 participants over time across the state. In terms of scope, we study component-specific costs of the program, differentiating between expenditures related to key program activities (community mobilization and financial inclusion), and expenditures related to overall operations and management. We also study the changes in total program costs with respect to core program activities, including the federation of SHGs, such as the formation of SHGs, VOs, CLFs, and bank linkages.

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.[9] From *Jeevika*'s expenditure components discussed earlier, we consider three expenditure components – Community Institution Development, Special Technical Assistance Fund, and Project Management. We do not include Community Investment Funds when studying the relationship of costs with scale because it is disbursed to

[7] The rest of this paper uses the terms "costs" and "expenditures" interchangeably.
[8] 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.
[9] This also implies that program attrition is not considered when estimating annual per member costs.

SHGs or their federations as a one-time grant or capital based on need and is usually a fixed amount (of Rupees 66,000, on average, [Kochar et al., 2020]).[10]

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 were eligible to receive *Jeevika*'s support on different activities in every year, we estimate two average cost models based on assumptions on the size of the target groups for these activities – (1) Model 1 – Cumulative Participation, under which 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 in each year among cumulative number of women mobilized into SHGs up to that year; and (2) Model 2 – Fixed-Period Participation, under which 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. In our conversations with the program implementation team, we were informed that *Jeevika* continues to support participants throughout their membership, and therefore the first assumption is more plausible. The nature of support is likely to vary over time, however, with older groups receiving more support with economic livelihoods and newer groups receiving support with bank and credit linkage.

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

$$AC_{it} = \frac{Annual\ expenditure\ on\ component\ i\ in\ year\ t}{Target\ number\ of\ women\ under\ component\ i\ in\ year\ t} \quad (1)$$

Here the target number of women is estimated separately for two models based on the two assumptions defined above. We extracted all program expenditure data in local currency (Indian Rupee) and converted to 2018 USD by first converting annual expenses to 2018 prices by

[10] Community Investment Funds are still considered a cost, but we do not hypothesize economies of scale for this cost category.

adjusting for Rupee inflation using the Consumer Price Index (CPI) method, and then converting to 2018 USD by adjusting for 2018 market exchange rate.[11]

We used the production function to test the possibility of increasing, decreasing or constant returns to scale, for the same program over time. We estimate the time-varying average cost function as a function of number of individuals served in first linear, and then quadratic form as shown in equations (2) and (3):

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

$$AC_t = \alpha_0 + \alpha_1 n_t + \alpha_2 n_t^2 + \epsilon_t \quad (3)$$

Here AC_t is the average program cost in year t (expressed in 2018 USD), and n_t is the 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. Our regressions cover twelve years of data, which is indisputably a small sample for a formal regression analysis. Yet, costing analyses rarely include data observed at more than one time-point because of data limitations. The goal of our analysis is to provide insight into how costs vary as program scale varies, which researchers have earlier carried out with much smaller sample sizes (see for example Tulloch, 2019).

Next, we investigate how total program expenditure changes with respect to specific economic activities. We use four different indicators of economic activities under *Jeevika*, focusing on primary program outcomes – (1) annual SHG formation; (2) annual Village Organization formation; (3) annual Cluster Level Federation formation; and (4) annual number of SHGs linked to the bank. The first three activities directly address the key output of *Jeevika* – formation of women-led community institutions at different levels. The fourth activity denotes the financial inclusion goal of the program. However, our data do not allow for disaggregating costs beyond the three categories of CID, PM, and STAF, implying that we cannot reliably estimate the cost of each specific activity.[12] Therefore, instead of average cost per unit of outcome achieved, we study how the total annual expenditure of program changes with respect to each of the four indicators, following equation (4).

[11] The CPI method estimates $Expenditure_{i2018} = Expenditure_{it} \times \frac{CPI_{2018}}{CPI_t}$

[12] As a result, we also face challenges in distinguishing economies of scale caused by group-based delivery models and economies of scale associated with other *Jeevika* program features.

$$Log(TC)_t = \alpha_0 + \alpha_1 Log(X_t) + \alpha_2 Log(X_t^2) + \epsilon_t \quad (4)$$

We estimate four different specifications of equation (4) for each economic activity as input for X_t (as opposed to one regression with the four activities because of high correlations across the activities). Because the four activities have very different scale of operations, we use log of the number of activities. If $\alpha_1 < 1$, we find evidence for economies of scale. We compare α_1 across the four activities to understand differences in economies of scale across economic activities.[13]

Finally, we estimate program cost-effectiveness during the pilot and scaled-up phases using findings on impact from existing studies (Datta, 2015; Hoffman et al., 2018; 2021). We primarily focus on the most direct source of program impact by examining household-level 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. Studies have shown that *Jeevika* resulted in a statistically significant and sizeable reduction in interest rates on loans taken by households, primarily due to reduced dependence on informal moneylenders (Hoffman et al., 2021).[14] Although we consider financial inclusion as the primary outcome, the program could create downstream effects on economic outcomes, such as consumption and productive asset ownership, as well as women's empowerment. Therefore, the cost-effectiveness for financial inclusion may present an incomplete picture of the overall cost-effectiveness of the program. To address this concern, we also discuss the program effects and cost-effectiveness for empowerment outcomes and asset ownership, including heterogeneous effects on asset ownership for landless households. However, the data do not allow for a comparative analysis for these outcomes at pilot versus scaled-up phases, because of the differences in their measurement across the two impact evaluations, and because the pilot phase does not report heterogeneous effects for landless households. Cost-effectiveness ratios were estimated by dividing the costs during program participation years by program impacts.

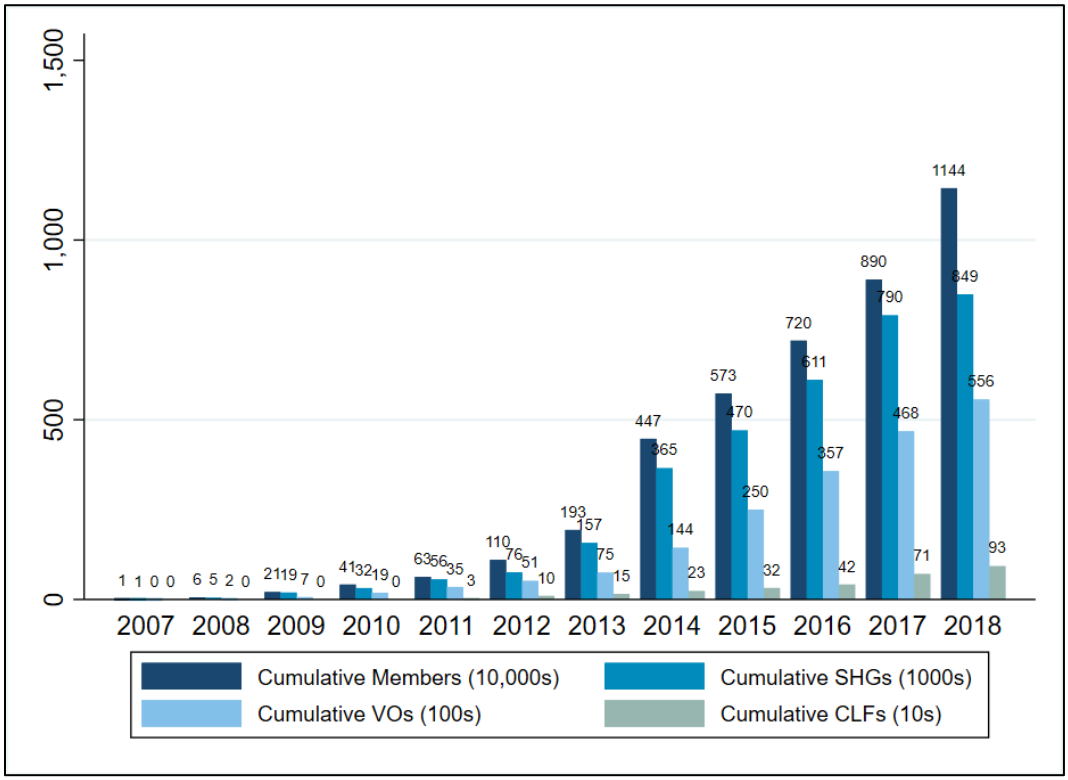
[13] Other studies focused on changes in the costs of pension funds with respect to scale using similar approaches (Bikker et al., 2012).
[14] Kochar et al. (2020) conducted a follow-up study to determine longer-term effects of the *Jeevika* program. We will discuss these results in a full holistic cost-effectiveness analysis of the *Jeevika* program.

Results

Expansion in Scale of Activities

We begin the analysis with an overview of the scale of *Jeevika*'s operations between 2007-08 to 2018-19. Figure 1 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 1, the pace of mobilization increased steadily post expansion, especially after 2012-13, and by 2018-19 over 11.4 million women had been mobilized into almost half a million SHGs.

Figure 1. *Jeevika* Program Scale-Up



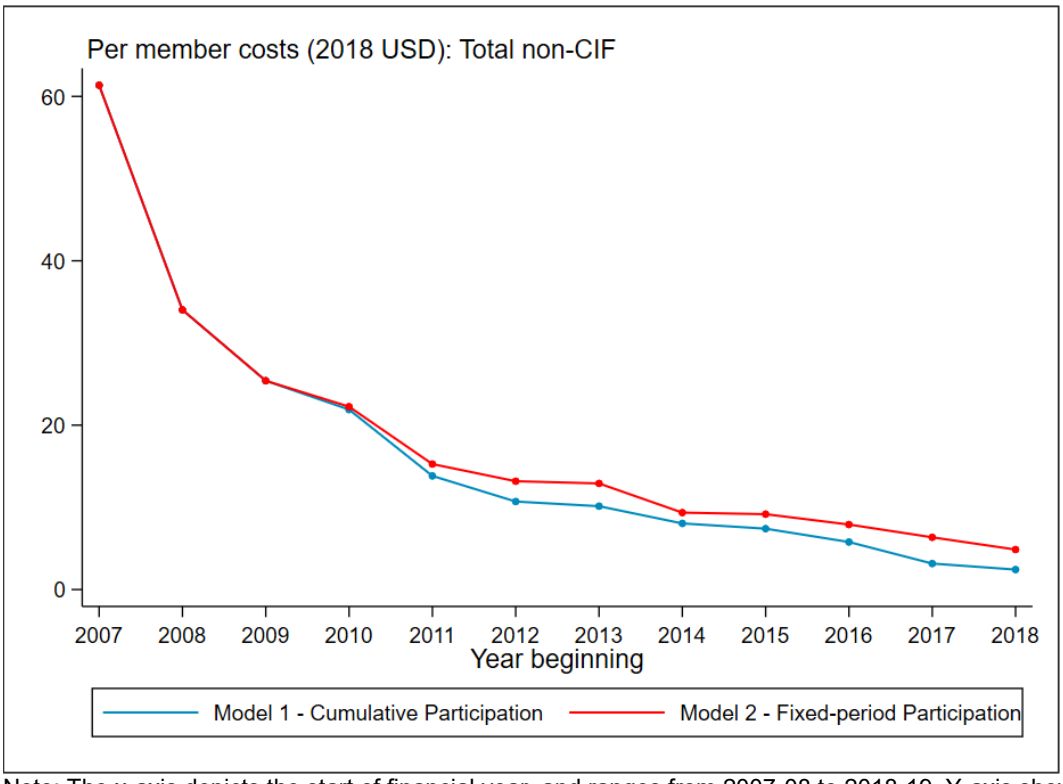
Note: The x-axis depicts the start of financial year, and ranges from 2007-08 to 2018-19.

Operating Expenditures

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 2 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 the year of initial participation (Model 2). By 2018-19, with a membership of 11.4 million beneficiaries, the per member expenditure ranged between $2.5 in Model 1 and $4.9 in Model 2.

Figure 2. *Jeevika* Annual non-CIF Expenditure Over Time



Note: The x-axis depicts the start of financial year, and ranges from 2007-08 to 2018-19. Y-axis shows per member combined annual expenditure on Community Institutional Development, Special Technical Assistance Fund, and Project Management. Model 1 assumes that expenditures are divided across all members irrespective of when they joined the program. Model 2 assumes that all members are supported for three years after joining the program.

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 $0.28 decline in expenditures per capita in Model 1, and a $0.49 decline in expenditures per capita 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 $31.0 in Model 1 to $35.6 in Model 2. Beyond that, increasing scale by an additional 100,000 members is associated with a decrease in per capita expenditure by 77 cents to $1.8. 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. This is likely because 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 11.4 million. Our regression estimates suggest that annual per capita expenditure for this level of outreach was $6.70 under Model 1, and $12.30 under Model 2.

Table 1. *Jeevika* Total Annual non-CIF Expenditure

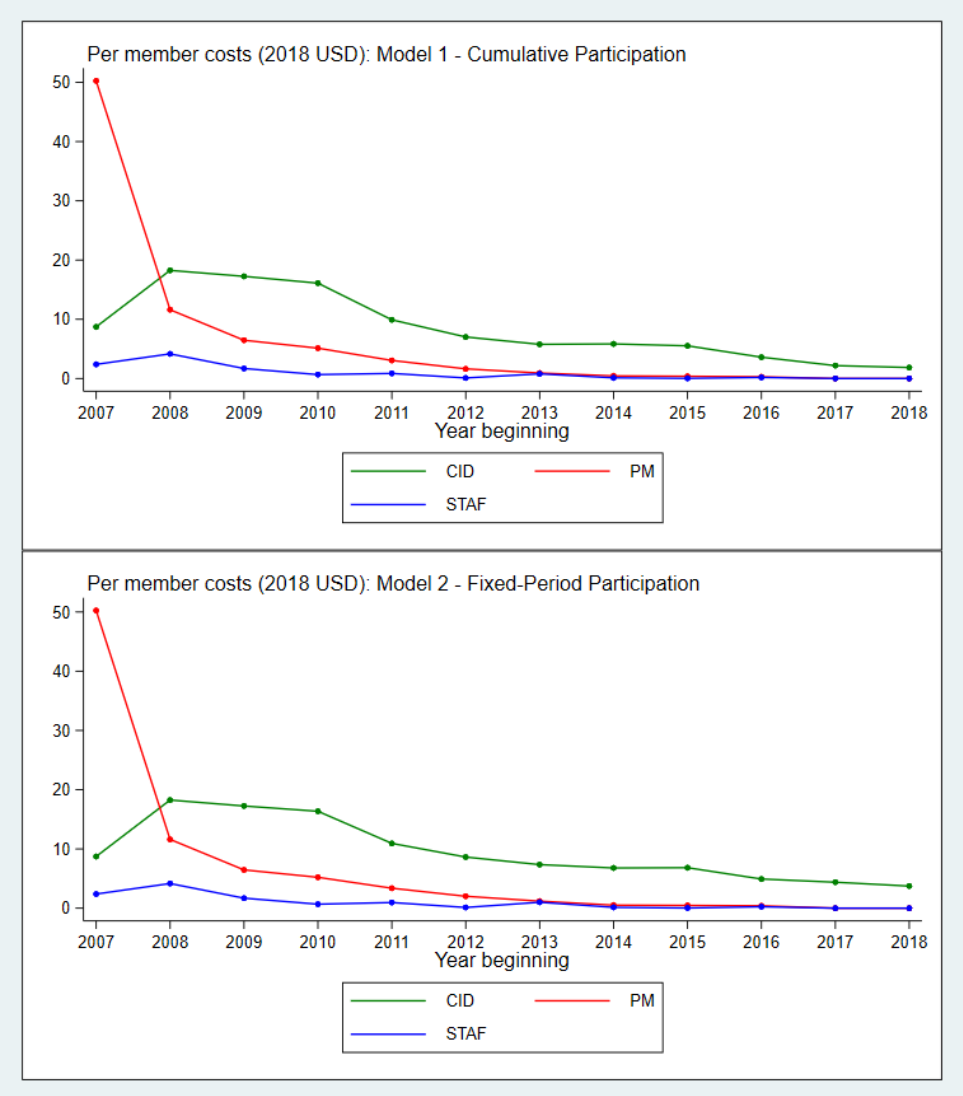
	Model 1 Cumulative participation		Model 2 Fixed-period participation	
Reach (in 100,000)	-0.280** (0.093)	-0.772** (0.299)	-0.489** (0.157)	-1.821* (0.812)
Reach (in 100,000) squared		0.005* (0.002)		0.025* (0.013)
Constant	26.835*** (6.735)	31.043*** (7.900)	29.730*** (6.941)	35.594*** (8.401)
Includes quadratic reach	N	Y	N	Y
Observations	12	12	12	12
R-squared	0.429	0.546	0.479	0.620
F test:	9.027	5.501	9.747	5.471

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Model 1 assumes that expenditures are divided across all members irrespective of when they joined the program. Model 2 assumes that all members are supported for three years after joining the program. Breusch-Godfrey test statistics for serial autocorrelation: (1) Cumulative participation, complete model with quadratic reach: Chi^2 value = 1.39; p-value = 0.24; (2) Fixed-period participation, complete model with quadratic reach: Chi^2 value = 0.55; p-value = 0.46. Both models fail to reject the null hypothesis of no serial correlation.

Next, we break down the overall non-CIF expenditure into its subcomponents – specifically, CID, STAF, and PM related expenditure. As shown in Figure 3, CID expenditure per member was highest in initial years (at $17.35 per capita), kept declining after 2009, and the decline

tapered off after 2012. At the maximum program outreach of 11.4 million members, CID expenditure was between $1.8 (under Model 1) and $3.7 (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 costs changed from $3.95 per capita when the project had an outreach of 59,000 members to 17 cents per member when the project had the highest outreach of 11.4 million members. Average expenditure related to PM activities also declined rapidly over time. In the first year of the project, PM expenditure was estimated at $47 per capita, which declined to $0.01 under both Models 1 and 2 by the latest year (2018-19).

Figure 3. *Jeevika* Non-CIF Expenditure Components Over Time



Note: The x-axis depicts the start of financial year, and ranges from 2007-08 to 2018-19. Y-axis shows per member annual expenditure by expenditure component. CID: Community Institutional Development; STAF: Special Technical Assistance Fund: PM: Project Management. Model 1 assumes that expenditures are divided across all members irrespective of when they joined the program. Model 2 assumes that all members are supported for three years after joining the program.

The CID component includes expenditures on project start-up, community mobilization, formation of higher-level federations, capacity management, and implementation of inclusive strategies to ensure that the project activities prioritize 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 reported to have spent lower amounts on the STAF component than initially estimated because of leveraging already established innovation linkages (World Bank, 2017). 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 setting up of the project Management Information System.

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. Increasing outreach by an additional 100,000 members is associated with the largest decline in PM-related expenditure (by \$0.15 to \$0.30; $p>0.10$), followed by CID expenditure (\$0.11 to \$0.18; $p<0.01$), while decline in STAF expenditure is slowest with respect to scale (by \$0.03 to \$0.04; $p<0.05$). We see that scale squared is positively associated with average expenditures per member, but this coefficient is not statistically significant in most specifications, 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 \$13.6 to \$14.6 per program participant in CID expenses, \$1.9 to \$2.2 per program participant in STAF expenses, and \$13.7 to \$18.9 per program participant in PM expenses. The same models predict that when the program operates at maximum scale, the per capita expenses reduce to \$2.2 to \$5.1 per program participant for CID, \$0.19 to \$0.48 per program participant for STAF, and \$2.3 to \$5.7 per program participant for PM related components.

Table 2. *Jeevika* Annual Non-CIF Expenditure Components: CID

	Model 1 Cumulative participation		Model 2 Fixed-period participation	
Reach (in 100,000)	-0.114***	-0.244**	-0.179***	-0.399
	(0.025)	(0.076)	(0.039)	(0.224)
Reach (in 100,000) squared		0.001*		0.004
		(0.001)		(0.004)
Constant	12.504***	13.619***	13.618***	14.588***
	(1.818)	(2.098)	(1.780)	(2.399)
Includes quadratic reach	N	Y	N	Y
Observations	12	12	12	12
R-squared	0.615	0.686	0.631	0.669
F test:	20.64	14.92	20.57	10.83

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Model 1 assumes that expenditures are divided across all members irrespective of when they joined the program. Model 2 assumes that all members are supported for three years after joining the program. Breusch-Godfrey test statistics for serial autocorrelation: (1) Cumulative participation, complete model with quadratic reach: Chi^2 value = 0.70; p-value = 0.40; (2) Fixed-period participation, complete model with quadratic reach: Chi^2 value = 0.08; p-value = 0.77. Both models fail to reject the null hypothesis of no serial correlation.

Table 3. *Jeevika* Annual Non-CIF Expenditure Components: STAF

	Model 1 Cumulative participation		Model 2 Fixed-period participation	
Reach (in 100,000)	-0.030**	-0.115**	-0.040**	-0.175*
	(0.012)	(0.048)	(0.015)	(0.079)
Reach (in 100,000) squared		0.001*		0.003
		(0.001)		(0.001)
Constant	1.751**	2.198**	1.834**	2.381***
	(0.571)	(0.655)	(0.570)	(0.668)
Includes quadratic reach	N	Y	N	Y
Observations	10	10	10	10
R-squared	0.360	0.531	0.385	0.565
F test:	6.538	4.307	7.479	4.927

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Model 1 assumes that expenditures are divided across all members irrespective of when they joined the program. Model 2 assumes that all members are supported for three years after joining the program. STAF expenditure is not reported for years 2017-18 and 2018-19. Breusch-Godfrey test statistics for serial autocorrelation: (1) Cumulative participation, complete model with quadratic reach: Chi^2 value = 0.41; p-value = 0.52; (2) Fixed-period participation, complete model with quadratic reach: Chi^2 value = 0.07; p-value = 0.79. Both models fail to reject the null hypothesis of no serial correlation.

Table 4. *Jeevika* Annual Non-CIF Expenditure Components: PM

	Model 1 Cumulative participation		Model 2 Fixed-period participation	
Reach (in 100,000)	-0.153	-0.528	-0.302	-1.471
	(0.094)	(0.331)	(0.173)	(0.937)
Reach (in 100,000) squared		0.004		0.022
		(0.002)		(0.015)
Constant	12.064	15.274	13.714	18.861*
	(6.936)	(8.800)	(7.619)	(10.233)
Includes quadratic reach	N	Y	N	Y
Observations	12	12	12	12
R-squared	0.184	0.281	0.235	0.374
F test:	2.689	1.418	3.026	1.598

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Model 1 assumes that expenditures are divided across all members irrespective of when they joined the program. Model 2 assumes that all members are supported for three years after joining the program. Breusch-Godfrey test statistics for serial autocorrelation: (1) Cumulative participation, complete model with quadratic reach: Chi^2 value = 0.15; p-value = 0.70; (2) Fixed-period participation, complete model with quadratic reach: Chi^2 value = 0.01; p-value = 0.93. Both models fail to reject the null hypothesis of no serial correlation.

Economies of Scale of Different Program Activities

We next analyze how annual program expenditures change in response to changes in annual (1) SHG formation, (2) Village Organization formation, (3) cluster-level federation formation, and (4) the number of SHGs linked to the bank. Table 5 shows that a 1% increase in each of the four activities considered is associated with a less than 1% increase in annual program expenditure – indicating economies of scale across the board. Among the four activities, number of CLFs formed is associated with largest gains, and SHG formation is associated with the least gains in scale economies. Specifically, annual program expenditure increased by 0.49% with a 1% increase in the number of annual CLFs formed, and by 0.77% with a 1% increase in the number of new SHGs formed. This result may reflect the fact that SHG formation is the first program activity as part of community mobilization, while CLF formation occurs at later stages, once groups have been functioning for a period of time, have formed Village Organizations, and the required capacity and infrastructure are established. Interestingly, we also find a high degree of economies of scope for the number of SHGs that are linked to formal credit.

Table 5. Logged Total Annual Expenditure as a Function of Different Economic Activities

	(1) Total cost (2018 USD) Logged	(2) Total cost (2018 USD) Logged	(3) Total cost (2018 USD) Logged	(4) Total cost (2018 USD) Logged
Logged SHGs formed	0.769***			
	(0.049)			
Logged SHGs linked to credit		0.542***		
		(0.048)		
Logged VOs formed			0.709***	
			(0.033)	
Logged CLFs formed				0.495***
				(0.138)
Constant	8.311***	11.004***	10.935***	14.743***
	(0.492)	(0.445)	(0.239)	(0.616)
Observations	12	12	12	12
R-squared	0.947	0.940	0.977	0.667
F test:	242.6	129.5	457.8	12.83

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The table includes estimates from four different regressions, each with the same outcome (total annual cost, logged) and different independent variables for each of the four economic activities.

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

Our findings suggest that large scale implementation of women's SHG programs has considerable potential to benefit from reduced costs due to economies of scale. 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 2021, covers over eleven 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 scale (Mansuri and Rao, 2012). Findings from two impact evaluations (Datta. 2015; Hoffman et al., 2021) – the first focusing on the first pilot phase of the project (pre-2012) and the latter on the second phase during scale-up (2012-2014) – indicated

that while the project was able to generate strong positive effects, especially on social empowerment, in the pilot phase, the second phase (post-2012) failed to see most of these individual-level and household-level effects.

An ethnographic study attributed the failure to produce similar effects in the second phase to the pressure to scale up quickly (Majumdar et al., 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.

We next 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.[15] 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 et al., 2017; Datta, 2015; Hoffman et al., 2018).[16] As shown in Table 6, we find an interesting pattern in the cost-effectiveness ratios (CERs) for high cost debt. Before 2012, that is in phase 1 of the project, the average high-cost loan amount decreased by $89 because of the *Jeevika* program, more than twice the average impact of $33 in phase 2. However, it costed 91 cents to reduce each additional dollar of high-cost debt in phase 1 – only slightly different from 88 cents in phase 2. 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. While program impact declined by almost 62% after it scaled up, it was 3 cents less costly for the program to achieve an additional dollar of benefit in reducing high-cost debts.

In line with Evans & Popova (2016), we also present confidence intervals for additional benefit per dollar spent on *Jeevika* to assess the uncertainty in cost-effectiveness estimates due to imprecision in estimation of impact. Figure 4 shows that every dollar spent on the program

[15] *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.

[16] 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.

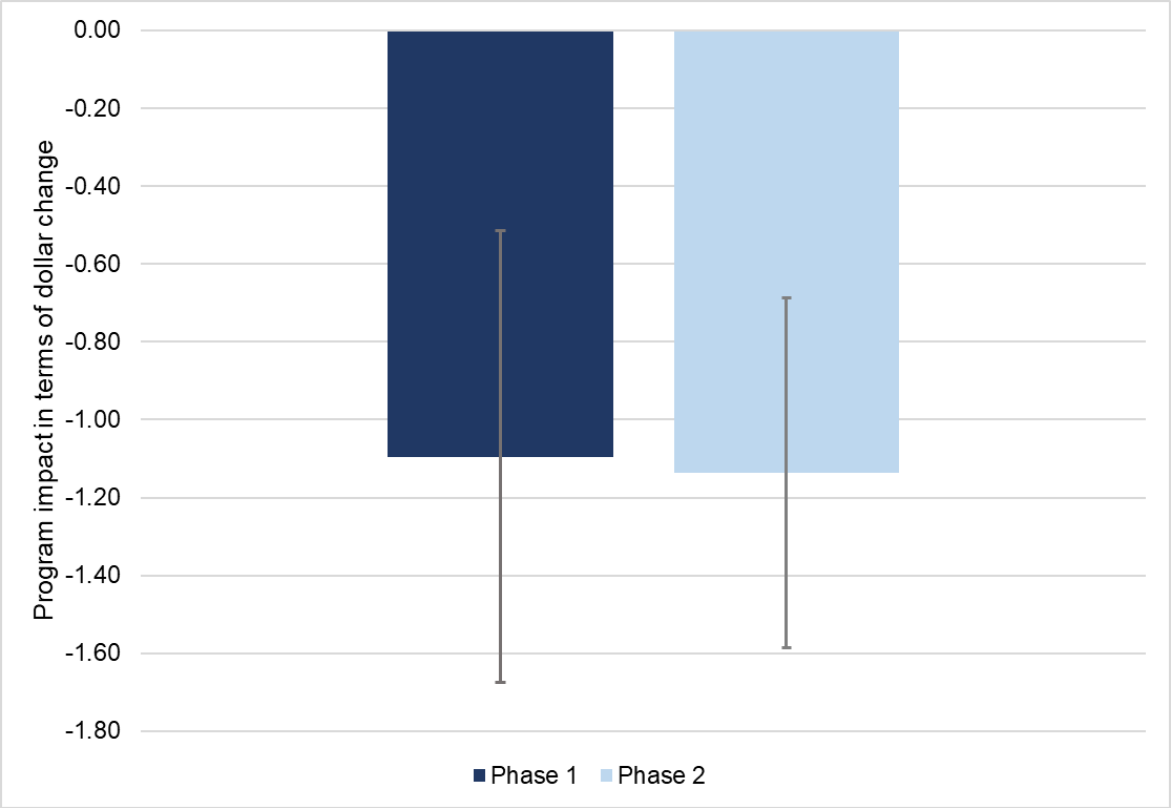
reduced high-cost loan amount by $1.10 in Phase 1, and by $1.14 in Phase 2. The 95% confidence intervals show that the lower bound is only slightly higher in Phase 2 while the upper bound is slightly higher in Phase 1. Overall, we still do not find any meaningful difference in the program cost-effectiveness over the two phases after considering the sensitivity of our estimates to uncertainty in the two impact estimates.

Table 6. Cost Per Dollar Change in Outcomes

	Phase 1	Phase 2
Primary costs, outcomes, and cost-effectiveness		
Average change in high-cost loan amount per household	89.08	32.83
Operating costs per capita	81.33	28.90
Cost per dollar of decrease in high-cost loan	0.91	0.88
Empowerment impacts and cost-effectiveness in Phase 1		
Cost per percentage-point increase in women attending Panchayat meetings	40.67	No significant impact on women's empowerment index
Cost per percentage-point increase in women participating in health decisions	27.11	
Cost per percentage-point increase in women visiting health center	27.11	
Cost per percentage-point increase in women having an opinion on politics	20.33	
Cost per percentage-point increase in women providing input on children's education	81.33	
Asset ownership impacts and cost-effectiveness for landless households in Phase 2		
Costs per SD increase in consumption asset index	Not measured	160.57
Costs per SD increase in productive asset index		722.56

Note: All costs and high-cost loan amounts 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. (2018) and Datta et al. (2015). SD=standard deviation.

Figure 4. Change in High-Cost Loan Amount for Every Dollar Spent on the Program



Note: All estimates are in 2018 USD. Error bars represent 95% confidence intervals. Costs are based on Model 1 assumptions –expenditures are divided across all members irrespective of when they joined the program.

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, while the phase 2 evaluation used a randomized controlled trial. 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.[17] 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.[18] 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

[17] 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.
[18] We will present these benefit-cost ratios in a forthcoming study that includes longer-term impact estimates of the Jeevika program.

action; while the Phase 2 evaluation showed only weak and mostly conflicting effects on empowerment indicators.

However, the Phase 2 evaluation did show significant negative effects on interest rates charged by informal lenders, which resulted in a reduction in the difference in average borrowing rates faced by landholding and landless households by two-thirds. The reduction in informal interest rates and the shift towards loans from formal sources both contributed to reducing the costs of borrowing, especially for landless households that had higher costs of borrowing at baseline. These impacts potentially led to other effects – the study showed statistically significant positive impacts on consumption asset and productive asset indices for landless households, despite very modest average treatment effects on consumption assets and null average treatment effects on productive assets for the full sample (Hoffman et al., 2018). Considering the wide reach of informal moneylenders among non-program participants, aggregate program benefits may increase even more following the program scale-up.

Current estimates of the cost-effectiveness of the scaled-up program for landless households indicate that it would cost $161 to increase the consumption asset index by one standard deviation, and $723 to increase the productive asset index by one standard deviation (Table 6). The reduced interest rates thus also seem to result in cost-effectiveness for other economic outcomes, at least for marginalized populations. However, the scaled-up program did not seem to achieve cost-effective outcomes for improving consumption and women's empowerment – neither for the full sample or for landless households, since the effects were negligible and insignificant for these outcomes. At the same time, we find evidence that the pilot phase resulted in a larger cost-effectiveness of *Jeevika* in achieving positive effects women's empowerment outcomes (Table 6), indicating that prioritizing program elements that led to positive effects on asset ownership and women's empowerment before the scaling may result in more positive cost-effectiveness for these outcomes even when costs are higher.

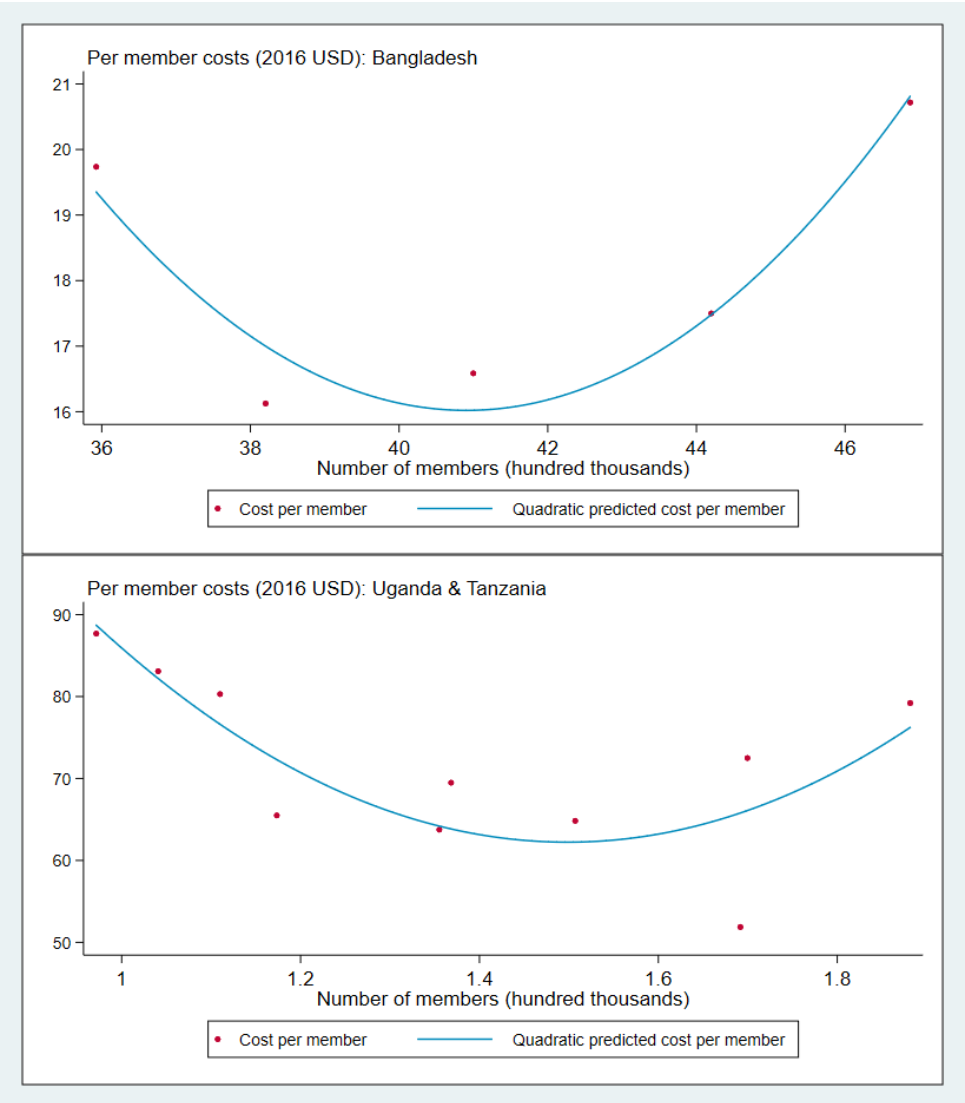
External Validity on Cost-Efficiencies

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, n.d.). 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 5 shows the operational costs per borrower with respect to total number of borrowers in the three countries. Between 2013 and 2017, 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.[19]

[19] 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, n.d.).

Figure 5. BRAC Microfinance Costs: 2013 to 2017



Note: The chart shows per member annual expenditure on y-axis, and number of members on x-axis. Bangladesh is shown separately from Uganda and Tanzania because of the wide disparity in scale. Scatter plots with polynomial fit lines are shown in both charts.

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 statistically 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 & Policy Implications

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-effectiveness and cost-benefit analyses into impact evaluations (Brown & Tanner, 2019). Second, while many development interventions often show positive impacts of pilot programs, 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 benefit from economies of scale in terms of lower costs, a factor that is especially crucial for group-based programs like SHGs which are able to deliver services to many people at once.

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 annual per capita expenditures on basic program activities (all non-CIF components) declined by 77 cents to $1.8 for every 100,000 new households, and the decline slowed down with a large increase in scale. Upon combining these cost estimates with impact estimates from the pilot phase and the scaled-up phase, we found that average high-cost loan amounts decreased by $89 in phase 1 during the pilot, more than twice the average impact of $33 in phase 2 during the program scale-up. However, this difference in program impact did not translate to large differences in cost-effectiveness, because the lower implementation costs of the scaled-up model were able to fully offset the lower benefits. Additional analyses of the operational costs of BRAC's group-based microfinance program in Bangladesh, Uganda, and Tanzania showed 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 often not available to operate programs at scale. Indeed, the expansion of *Jeevika*'s SHG program faced similar challenges. Along with the implementation challenges noted earlier, *Jeevika* 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, including 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 during the scale-up as during the pilot of the program.

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. These 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 trends may be very different for other women's group programs with economic objectives or more generally for international development programs.

For SHGs in particular it is critical to examine how changes in costs after scale-up are related to the group-based nature of SHGs. A recent study by Grantham-McGregor et al. (2020) suggests equally positive effects of home visits and group sessions on early childhood cognition in Odisha, India, but group sessions were 3.5 times less expensive than home visits. This finding indicates that training SHG members in a group setting could lead to more cost-effective outcomes than training SHG members individually, and differences in the cost-effectiveness of training of groups versus individuals may be even more pronounced after the scale-up of SHG programs. *Jeevika* data on specific cost categories did not include sufficient detail about group-based versus individual activities, but future studies should examine this question by collecting primary data on cost ingredients by economic activity, following existing guidelines and tools.[20]

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 biased view about the possibility of achieving positive impacts at scale when studies fail to account for the existence of economies of 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. Importantly, however, examining the cost-effectiveness of programs is less meaningful if programs fail to produce positive impacts once scaled-up. While *Jeevika* did have positive effects on consumption and productive asset ownership for landless households after scaling up, the program had only modest average treatment effects on consumption asset ownership and no positive effects on productive asset ownership or women's decision-making power (Hoffmann et al., 2021). Our findings show that the reduced impact came alongside significant cost reductions. We show that these cost reductions are related to scale economies, while previous qualitative research indicates that reduced impacts may be related to changes in implementation quality after scale-up. For example, studies showed that the program was less consistent in applying some of the ritualized processes of group activities after the scale-up, and

[20] See, for example, Bhula & Murphy (2020) and Siwach et al. (2019).

coordinators spent less time on gathering contextual information and adapting their messaging for mobilization (Hoffman et al., 2018; Majumdar et al., 2017).

Our findings combined with findings from a recent study of the NRLM (Kochar et al., 2020) further indicate that SHGs that operate under a federated structure are possibly more cost-effective than groups that are more disjoint. *Jeevika*, and the NRLM program more broadly, hold federations at the center of their SHG programming. Our findings indicate that the formation of Cluster-Level Federations in later stages of the SHG programming comes with lower marginal costs than the formation of SHGs. Specifically, we found a 0.49% increase in annual costs for a 1% increase in the number of Cluster-level Federations formed, signifying large potential for scale economies from the formation of federations. Evidence from other studies shows marked improvements in the performance of SHGs after the formation of Village Organizations and Cluster-level federations. Specifically, Kochar et al. (2020) find that linking SHGs to Village Organizations and Cluster Level Federations is associated with improved financial access and use of funds.

From a policy perspective, these findings suggest that SHG programming at scale could increase its cost-effectiveness by (1) prioritizing program elements that evaluations of pilot programs identify as critical for achieving positive impacts on women's empowerment and asset ownership, and (2) by continuing to place a large emphasis on the formation of higher-level federations. Qualitative evidence indicates that gathering contextual information and adapting messaging for mobilization was critical for achieving positive effects on women's empowerment. Strengthening these program elements at scale could increase the cost-effectiveness of *Jeevika* even with an increase in costs, as long as *Jeevika* maintains the elements that were critical for success during the pilot phase. Further, the costs of contextual information and mobilization are mostly one-time costs, which likely do not require long-term investments. Our results further suggest low marginal costs of the formation of federations, and Kochar et al. (2020) suggest that the formation of federation can increase the effectiveness of the NRLM. Investments in the formation of higher-level federations are thus likely cost-effective, especially because federations serve as primary platforms for creating women-led community institutions that can leverage community resources and perform negotiations for the collective economic upliftment of their SHGs and its members.

Finally, 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, 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. The average cost per beneficiary at scale will likely decrease further if the size of beneficiaries goes beyond the intended target participants. Future research should estimate these long-term general equilibrium effects, and also include estimations of 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

Abramsky, T., Devries, K., Kiss, L., Francisco, L., Nakuti, J., Musuya, T., Kyegombe, N., Starmann, E., Kaye, D., Michau, L., & Watts, C. (2012). A community mobilisation intervention to prevent violence against women and reduce HIV/AIDS risk in Kampala, Uganda (the SASA! Study): study protocol for a cluster randomised controlled trial. *Trials*, *13*(1), 1-22.

Anderson, C. L., de Hoop, T., Desai, S., Siwach, G., Meysonnat, A., Gupta, R., Haroon, N., Howlett, M., Kolla, N., Sidhu, A., Paul, S., Belyakova, Y., & Singh, R. S. (2019). *Investing in women's groups: A portfolio evaluation of the Bill & Melinda Gates Foundation's investments in South Asia and Africa*. Research Brief. Retrieved from http://www.womensgroupevidence.org . Accessed November 01, 2020.

Artuc, E., Cull, R., Dasgupta, S., Fattal, R., Filmer, D., Giné, X., Jacoby, H.G., Jolliffe, D., Kee, H.L., Klapper, L., & Kraay, A. (2020). Toward Successful Development Policies: Insights from Research in Development Economics. *World Bank Policy Research Working Paper*, (9133).

Bhula, R., Mahoney, M., & Murphy, K. (2020). Conducting cost-effectiveness analysis (CEA). J-PAL Research Resources. Retrieved from https://www.povertyactionlab.org/resource/conducting-cost-effectiveness-analysis-cea . Accessed April 01, 2021.

Bijlmakers, L., Cornelissen, D., Cheelo, M., Nthele, M., Kachimba, J., Broekhuizen, H., Gajewski, J. & Brugha, R. (2018). The cost of providing and scaling up surgery: a comparison of a district hospital and a referral hospital in Zambia. *Health policy and planning*, *33*(10), 1055-1064.

Bikker, J. A., Steenbeek, O. W., & Torracchi, F. (2012). The Impact of Scale, Complexity, and Service Quality on the Administrative Costs of Pension Funds: A Cross-Country Comparison. *Journal of Risk and Insurance*, *79*(2), 477-514.

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 . Accessed November 01, 2020.

Breza, E., & Kinnan, C. (2018). Measuring the equilibrium impacts of credit: Evidence from the Indian microfinance crisis. *National Bureau of Economic Research Working Paper*, (w24329).

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.

Brooker, S., Kabatereine, N. B., Fleming, F., & Devlin, N. (2008). Cost and cost-effectiveness of nationwide school-based helminth control in Uganda: intra-country variation and effects of scaling-up. *Health policy and planning*, *23*(1), 24-35.

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

Bruns, B., Macdonald, I. H., & Schneider, B. R. (2019). The politics of quality reforms and the challenges for SDGs in education. *World Development*, *118*, 27-38.

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., Makwenda, C., Batura, N., Haghparast-Bidgoli, H., Hunter, R., & 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, U., & Rao, V. (2018a). 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, U., & Rao, V. (2018b). 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

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.

Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, *210*, 2-21.

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. Retrieved from https://faculty.wcas.northwestern.edu/~sjv340/womens_groups.pdf . Accessed April 01, 2021.

Evans, D. K., & Popova, A. (2016). Cost-effectiveness analysis in development: Accounting for local costs and noisy impacts. *World Development*, *77*, 262-276.

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.

Government of India. (2011). Census 2011 results. Retrieved from http://censusindia.gov.in/2011-prov-results/data_files/india/Final_PPT_2011chapter7.pdf?_sm_au_=iVVS1tPffSqnVskj . Accessed November 10, 2021.

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

Grantham-McGregor, S., Adya, A., Attanasio, O., Augsburg, B., Behrman, J., Caeyers, B., Day, M., Jervis, P., Kochar, R., Makkar, P., & Vats, K. (2020). Group sessions or home visits for early childhood development in India: a cluster RCT. *Pediatrics*, *146*(6).

Harper, M. (2002). Promotion of self help groups under the SHG bank linkage programme in India. In Seminar on SHG-bank linkage programme at New Delhi (pp. 25–26). Retrieved from https://pdfs.semanticscholar.org/37e9/60be6d6bea3fe4b 5e718cbf0f3aa80efeb5b.pdf . Accessed November 10, 2020.

Hoffman, V., Rao, V., Datta, U., Sanyal, P., Surendra, V., & Majumdar, S. (2018). Poverty and empowerment impacts of the Bihar Rural Livelihoods Project in India. International Initiative for Impact Evaluation. *Impact Evaluation Report* (71).

Hoffmann, V., Rao, V., Surendra, V., & Datta, U. (2021). Relief from usury: Impact of a self-help group lending program in rural India. *Journal of Development Economics*, *148*, 102567.

International Institute for Population Sciences (IIPS) and ICF. (2017). National Family Health Survey (NFHS-4), 2015-16: India. Mumbai: IIPS. Retrieved from https://dhsprogram.com/pubs/pdf/FR339/FR339.pdf . Accessed September 01, 2019.

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. Consultative Group to Assist the Poor (CGAP) *Occasional Paper* (No. 12).

Jan, S., Ferrari, G., Watts, C. H., Hargreaves, J. R., Kim, J.C., Phetla, G., Morison, L.A., Porter, J.D., Barnett, T., & Pronyk, P.M. (2011). Economic evaluation of a combined microfinance and gender training intervention for the prevention of intimate partner violence in rural South Africa. *Health policy and planning*, *26*(5), 366-372.

Jeevika. (n.d.) Jeevika Annual Reports 2006-07 to 2015-16. Retrieved from http://brlp.in/web/brlp/annual-audit-reports . Accessed February 01, 2019.

Jeuland, M., Pattanayak, S. K., Tan Soo, J. S., & Usmani, F. (2020). Preferences and the effectiveness of behavior-change interventions: Evidence from adoption of improved cookstoves in India. *Journal of the Association of Environmental and Resource Economists*, *7*(2), 305-343.

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

Karlan, D., Savonitto, B., Thuysbaert, B., & Udry, C. (2017). Impact of savings groups on the lives of the poor. *Proceedings of the National Academy of Sciences*, *114*(12), 3079-3084.

Kochar, A., Barooah, B., Jain, C., Closepet, N., Narayanan, R., Sarkar, R., & Shah, R. (2020). Impact Evaluation of the National Rural Livelihoods Programme. International Initiative for Impact Evaluation. *Impact Evaluation Report* (128).

Lipsky, M. (2010). *Street-level bureaucracy: Dilemmas of the individual in public services (Reprinted ed.)*. New York, NY: Russel Sage Foundation.

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. *World Bank Policy Research Working Paper*, (8039).

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. *World Bank Policy Research Working Paper*, (6239).

Ministry of Rural Development. (2011). Natural Rural Livelihoods Mission programme implementation plan. Retrieved from http://www.hprural.nic.in/livelihoods_mission.pdf . Accessed November 01, 2020.

Muralidharan, K., Niehaus, P., & Sukhtankar, S. (2020). Identity verification standards in welfare programs: experimental evidence from India. *National Bureau of Economic Research Working Paper* (26744).

Paul, B. V., Dutta, P. V., & Chaudhary, S. (2021). Assessing the Impact and Cost of Economic Inclusion Programs. *World Bank Policy Research Working Paper*, (9536).

Prennushi, G., & Gupta, A. (2014). Women's Empowerment and Socio- Economic Outcomes Impacts of the Andhra Pradesh Rural Poverty Reduction Program. *World Bank Policy Research Working Paper*, (6841).

Pressman, J. L., & Wildavsky, A. B. (1984). *Implementation: How great expectations in Washington are dashed in Oakland*. Berkeley, CA: University of California Press.

Prost, A., Colbourn, T., Seward, N., Azad, K., Coomarasamy, A., Copas, A., Houweling, T.A., Fottrell, E., Kuddus, A., Lewycka, S., & 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*.

Ricks, J. I., & Doner, R. F. (2021). Getting institutions right: Matching institutional capacities to developmental tasks. *World Development*, *139*, 105334.

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.

Siwach, G., de Hoop, T., Ferrari, G., & Belyakova, Y. (2019). Guidelines on Estimating the Costs and Cost-Effectiveness of Women's Groups in International Development. Evidence Consortium on Women's Groups. Retrieved from https://womensgroupevidence.org/our-work/guidance-for-cost-effectiveness . Accessed April 01, 2021.

Slater, R., S. Wiggins, L. Harman, M. Ulrichs, L. Scott, M. Knowles, P. Pozarny, and G. Calcagnini. (2016). *Strengthening Coherence between Agriculture and Social Protection: Synthesis of Seven Country Case Studies*. Rome: FAO.

Tankha, A. (2002). Tankha, A. (2002). Self-help groups as financial intermediaries in India: Cost of promotion, sustainability and impact. *New Delhi: Sa-Dhan, study prepared for ICCO and Cordaid, The Netherlands*.

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

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 from http://documents.worldbank.org/curated/en/675101468269084791/pdf/39039.pdf . Accessed February 01, 2019.

World Bank. (2012). *Implementation Status & Results: Bihar Rural Livelihoods Project – Jeevika.* Retrieved from http://documents.worldbank.org/curated/en/953391468771088485/pdf/ISR0Disclosabl0202020120134016973520 5.pdf . Accessed March 01, 2019.

World Bank. (2017). *A Decade of Rural Transformation: Lessons Learnt from the Bihar Rural Livelihoods Project – Jeevika.* Retrieved from http://documents.worldbank.org/curated/en/298391515516507115/pdf/122548-WP-P090764-PUBLIC-India-BRLP-Booklet-p.pdf . Accessed March 01, 2019.

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 . Accessed November 01, 2020.

World Bank. (2018b). *India State Briefs: Bihar.* Retrieved from https://www.worldbank.org/en/news/feature/2016/05/26/india-states-briefs. Accessed September 01, 2019.

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