NBER WORKING PAPER SERIES

RISK SHARING, COMMITMENT CONSTRAINTS AND SELF HELP GROUPS

Orazio Attanasio
Anjini Kochar
Aprajit Mahajan
Vaishnavi Surendra

Working Paper 31245
http://www.nber.org/papers/w31245

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2023, Revised June 2026

Evaluations of group savings and lending programs have largely focused on average impacts, rather than distributional impacts, finding modest effects on long-term economic well-being. Exploiting random variation in institutional capacity generated by the randomized roll-out of a self-help group lending program in rural Bihar, India (Hoffmann et al., 2021), this paper finds that risk sharing improved more where institutional capacity was greater. Building on models of informal insurance with limited commitment, the paper provides evidence of a specific channel: pre-existing program scale improves the quality and functioning of groups, which in turn raises the value of continued membership, relaxing the participation constraints that limit risk sharing.

Risk Sharing, Commitment Constraints and Self Help Groups
Orazio Attanasio, Anjini Kochar, Aprajit Mahajan, and Vaishnavi Surendra
NBER Working Paper No. 31245
May 2023, Revised June 2026
JEL No. D14, G21, I38, O13, O16

ABSTRACT

Evaluations of group savings and lending programs have largely focused on average impacts, rather than distributional impacts — finding modest effects on long-term economic well-being. In this paper, we exploit the randomized roll-out of a self-help group lending program in rural Bihar, India (Hoffmann et al., 2021) to demonstrate that well-functioning groups facilitate risk-sharing within rural communities. We find no impact of the program on risk-sharing, measured as a reduction in the variance of consumption growth, in the aggregate. However, the program significantly improves risk-sharing in regions where it had greater institutional capacity and was better implemented. Building on our theoretical framework, we provide evidence of a specific channel of impact: program quality and pre-existing scale improve the quality and functioning of groups, which in turn increase the insurance value of the program to communities.

Orazio Attanasio
Yale University
Department of Economics
and CEPR
and also NBER
orazio.attanasio@yale.edu

Anjini Kochar
International Initiative for
Impact Evaluation (3ie)
anjinikochar@gmail.com

Aprajit Mahajan
University of California, Berkeley
and NBER
aprajit@gmail.com

Vaishnavi Surendra
Precision Development (PxD)
vaishnavi.s@gmail.com

1 Introduction

A large literature, reviewed in Banerjee et al. (2015b), evaluates the impacts of savings and lending group programs, finding significant but often modest impacts on average outcomes such as income or consumption (Meager, 2019). However, two dimensions remain largely unexplored. First, such programs might have meaningful distributional impacts (despite small mean impacts)—in particular by improving risk-sharing and informal insurance of idiosyncratic shocks, whose welfare effects can be substantial in low-income contexts. Second, mean impacts aggregate across heterogeneous implementation settings: variable institutional capacity, for instance, may generate disparate impacts of a given program. While potentially important, such variation is rarely studied.

We explore both these dimensions by supplementing previous analyses of a large rural self-help group (SHG) program RCT in Bihar—*Jeevika*, evaluated by Hoffmann et al. (2021)—with alternative analytical measures and additional data. First, we measure risk sharing by focusing on the evolution of the cross-sectional distribution of household consumption, thereby overcoming the lack of information on income and income shocks that limits standard tests. Second, we collect information on pre-existing program scale as our measure of institutional capacity—which we view as a policy-induced source of treatment heterogeneity.

Jeevika is one of the world's largest rural SHG programs, covering 68% of Bihar's rural households by 2021 as part of India's National Rural Livelihoods Mission (NRLM). The randomized rollout (Hoffmann et al., 2021) spans 179 Gram Panchayats (GPs) (groups of 2–4 villages) nested within 16 administrative blocks. The randomization used a block-stratified matched-pair design with the GP as the unit of randomization and the program was implemented in all villages of a treated GP. A 2011 baseline survey was followed by program rollout in treatment GPs between January and April 2012 and a 2014 endline survey; control GPs received the program after the endline. We complement the RCT with two new data sources: a 2019 follow-up survey of study SHGs, providing measures of group functioning and quality, and NRLM's Management Information System (MIS), providing administrative records on SHG formation, membership, and location.

The program does not operate GP-by-GP in isolation: administrative staff and community cadres are allotted a geographic *operating area* covering multiple contiguous GPs within a block, and draw substantially on the pre-existing SHGs within that area for program implementation. Because pairs (within a block) are not matched by operating-area geography, the fraction of control GPs in any given operating area is effectively random, so the actual rolled-out program area (operating area minus control GP area) also varies randomly. We define the *initial program scale* as the ratio of pre-2010 SHGs to this rolled-out area, and use it as our measure of institutional capacity: higher values reflect more intensive engagement of fixed pre-existing resources with each treated village.

Establishing whether SHGs improve risk sharing requires a suitable measure. The standard approach, following Townsend (1994), examines the consumption responses to idiosyncratic income shocks, which would be absent in the presence of perfect risk sharing after controlling for changes in aggregate income in a risk-sharing group. However, reliable income and shock data are unavailable in our setting. Therefore, we follow Attanasio and Székely (2004) (who build on Deaton and Paxson, 1994; Albarran and Attanasio, 2003), exploiting a different implication of risk sharing. If log consumption can be interpreted as an approximation of log marginal utility, under perfect insurance the within-village variance of consumption growth should be zero.[1] Therefore, non-zero within-village variance of consumption growth indicates the absence of perfect risk sharing, and movements toward zero in such a measure indicate improvements in risk sharing.

We show that, while all treated areas experienced improvements in credit access, risk-sharing benefits were concentrated in higher-capacity areas. We find that a one-unit increase in institutional capacity reduces the variance of consumption growth by 7% in treated villages.

[1] Alternatively, the variance of log-consumption should be constant over time.

Consistent with models of informal insurance under limited commitment (Ligon et al., 2002), the evidence points to group quality as the operative mechanism: higher-capacity areas produced SHGs with larger accumulated funds and stronger governance, raising the value of continued membership and thereby tightening enforcement of informal insurance arrangements.

To interpret these findings, we further relate group quality and household responses to shocks to the program's institutional capacity. Using data from the 2019 SHG survey, we show that higher institutional capacity leads to better-functioning SHGs, with larger accumulated funds and stronger governance measured 5–7 years after group formation. Among treated villages, emergency borrowing conditional on reported shocks declines with institutional capacity, while credit and savings access do not vary. This pattern suggests that informal insurance substitutes for emergency credit, and households with better-functioning groups need not necessarily borrow to cope with shocks. This is an important and novel finding as it suggests that the improved risk sharing does not originate from the existence of a *savings fund per-se*, but by better functioning of informal loans and transfers.

We interpret these findings through models of informal insurance with limited commitment (e.g., Ligon et al., 2002). Within such models, better-functioning groups raise the value of membership, relaxing participation constraints to informal channels that limit risk sharing. In other words, the program fostered a small but cohesive network of women who support one another through mutual transfers, outside of formal loans from SHG funds. Such transfers are likely motivated by a desire to preserve group cohesion, given the substantial long-term benefits of participation in well-functioning SHGs.

This paper contributes to a growing literature on programs that operate at scale, where a central concern is whether capacity constraints adversely affect implementation and impacts (Banerjee et al., 2017a; Muralidharan and Niehaus, 2017; Muralidharan et al., 2023), and why program effects often diminish at scale (Vivalt, 2020; List, 2024). Existing studies identify several sources of implementation heterogeneity: government versus NGO delivery (Vivalt, 2020; Bold et al., 2018), delivery models, program take-up, program fidelity (Angrist and Meager, 2023), and input supply constraints at scale (Davis et al., 2017). We document a distinct source, i.e., planned variation in institutional capacity generated by the sequencing of program expansion. Rather than treating implementation capacity as fixed, phased expansion allows institutional learning and staff capacity to develop over time. This paper finds that group quality and functioning, a dimension of program fidelity, determine informal insurance benefits, with implications for how large-scale roll-outs are sequenced.

The paper also contributes to the large literature evaluating group savings and lending programs in developing countries. A number of randomized evaluations (Hoffmann et al., 2021; Angelucci et al., 2015; Attanasio et al., 2015; Augsburg et al., 2015; Banerjee et al., 2015a; Crépon et al., 2015; Tarozzi et al., 2015, reviewed in Banerjee et al. (2015b), meta-analysis in Meager (2019)), have found modest impacts on mean household outcomes that reflect long-term economic well-being, such as income and consumption, with more recent work showing effects are concentrated in the upper tail of profit and consumption distributions (Meager, 2022), and varying by household characteristics such as prior entrepreneurial experience (Banerjee et al., 2017b). Some studies of programs at scale find impacts on household consumption (Breza and Kinnan, 2021; Kaboski and Townsend, 2012).

A large body of work documents partial but significant informal insurance within villages (Townsend, 1994; Udry, 1994), and a related literature examines how financial interventions interact with risk sharing (Fischer, 2013; Attanasio et al., 2018). Yet this work provides little direct evidence on how group-based programs affect informal insurance, or on which aspects of group design explain their performance. With few exceptions (Demont, 2022), the impact of group formation on risk sharing remains virtually unstudied. Our access to SHG-level survey data and administrative records allows us to address this gap directly.

Our contribution relates most closely to the literature on participation constraints and risk sharing (Ligon et al., 2002). Informal insurance arrangements are sustained when members' value of continued participation exceeds their outside options. Understanding what determines these constraints is therefore central to understanding why informal insurance falls short of the full-risk-sharing benchmark. Feigenberg et al. (2013) study how microcredit program rules shape within-group risk sharing and repayment. Bold (2009) builds on Genicot and Ray (2003) to allow for endogenous group formation, showing that coalition-proof punishments can reduce enforcement severity by improving outside options. Dubois et al. (2008) show that formal contracting can relax binding participation constraints, though in their framework formality enables income smoothing through formal channels and thereby *reduces* the insurance value of group membership.

We identify a distinct and opposing channel. Administrative oversight and institutional capacity enhance group quality, raising the value of membership. Exclusion from a well-functioning self-help group is a harsher punishment than exclusion from a poorly functioning one. Higher group quality therefore *tightens* participation constraints and *strengthens* informal insurance—a mechanism that, to our knowledge, has not been formally studied.

The rest of the paper is organized as follows. We start in Section 2 with the sketch of a conceptual framework relating imperfect risk sharing to SHG quality. In Section 3, we provide information on the SHG program we consider, its different phases, and its institutional capacity. In Section 4, we describe the nature of the RCT that was designed to evaluate the program's impacts, and the plausibly exogenous variation in institutional capacity that it generated. Section 5 and Section 6 present our empirical results. Finally, Section 7 concludes the paper.

2 A Conceptual Framework for Risk Sharing and SHGs

We draw on the extensive literature on risk sharing in rural economies to measure within-village risk sharing and to develop a framework linking it to the presence and development of self-help groups (SHGs) in a village. A natural starting point is the benchmark of perfect within-group risk sharing in response to idiosyncratic shocks (Townsend, 1994). This approach allows us to analyze risk sharing *within* any arbitrarily defined group, while still permitting risk to be shared outside the group, focusing on the observed consumption allocations, without taking a stand on the mechanisms that generate them. Deviations from this benchmark provide a straightforward way to characterize observed allocations and quantify the degree of actual risk sharing.

Testing perfect risk sharing. Townsend (1994) considers a social planner who maximizes the expected utility of a risk sharing group, q. Within the group, each individual is assigned a *Pareto weight* in the planner's objective function. This *Pareto weight* can represent a variety of factors, such as an individual's wealth or position in the group. While some frictions, such as imperfectly enforceable contracts, can result in changes in the Pareto weights, under perfect risk sharing they are given, and the theory is silent about what determines them. The only constraint that the social planner faces is that of aggregate resources. Therefore, neglecting aggregate savings for notational simplicity,[2] the planner's problem can be written as choosing the group's consumption allocations, $c_t^{i,q}$, to maximize the following function.

$$W^q = \sum_{i=1}^{I_q} \lambda_i^q E_0 \sum_{t=0}^{\infty} \beta^t u(c_t^{i,q}) \tag{1}$$

$$subject \;\; to \quad \sum_i c_t^{i,q} \leq \sum_i y_t^{i,q}, \qquad \forall t.$$

The first order conditions for this problem's Lagrangian are given by the following equation:

$$\beta^t u'(c_t^{i,q})\lambda_i^q = \theta_t^q \qquad \forall i, t. \tag{2}$$

where θ_t^q is the (non-negative) Lagrange multiplier associated with the feasibility constraints

[2] Saving can be added to this problem, under the assumption and it is observable and contractable upon.

relevant in time, t.[3]

Assuming power utility and taking the log of the corresponding expression we get:

$$\ln c_t^{i,q} = \frac{1}{\gamma}(t \ln \beta - \ln \theta_t^q + \ln \lambda_i^q) \quad \forall i, t. \tag{3}$$

Taking first differences of equation (3) we get:

$$\Delta \ln c_t^{i,q} = \frac{1}{\gamma} \ln \beta - \frac{1}{\gamma} \Delta \ln \theta_t^q \quad \forall i, t. \tag{4}$$

Equations (3) and (4) capture the key implication of the perfect–risk-sharing model. The level equation implies that individual consumption is determined by an individual fixed effect, reflecting the person's *Pareto weight*, and a time fixed effect capturing aggregate resource constraints. The differenced equation shows that changes in individual marginal utility (here approximated by log consumption) are affected *only* by changes in aggregate resources.

The assumption of power utility, with Constant Relative Risk Aversion (CRRA), is central to deriving these equations that are linear in logs. However, alternative assumptions are possible. We discuss those in Appendix D.1. In our context, eqs. (3) and (4) can be interpreted as approximations to more nuanced utility specifications. These approximations could be particularly useful for our approach, which is based on (changes in the) cross sectional variances of log utility. Equations (3) and (4) form the basis for the tests in Townsend (1994): variables capturing idiosyncratic resources—such as individual income levels in eq. (3) or income changes in eq. (4)—should not enter significantly in these regressions. The magnitude and significance of their coefficients, therefore, provide a measure of departures from perfect risk sharing.

This approach is attractive because it relies only on the properties of resource allocation

[3]We note that these equations hold in any state of the world, so that we do not have an expected value. We are abusing notation here, in that the right-hand side of the f.o.c. should be multiplied by the probability of every state of the world considered. We can assume, without loss of generality, that such a term is absorbed by the multiplier θ_t^q.

within a group. The hypothesis being tested is whether a group (defined by the researcher) can, given its resource constraints, insure its members against a certain type of shock. It should be noted, however, that the empirical implementation of these tests requires information on both income and consumption at the individual level.

A different approach takes eqs. (3) and (4) in a complementary direction. Following Attanasio and Székely (2004), who build on an idea in Deaton and Paxson (1994), one can examine the within-group variance of the terms in these equations. Applied to eq. (3), this yields:

$$Var_q(\ln c_t^{i,q}) = \frac{1}{\gamma^2} Var_q(\ln \lambda_i^q) \quad \forall i, t. \tag{5}$$

This equation highlights a central implication of perfect risk sharing: within a risk–sharing group, the distribution of marginal utility remains constant and depends only on the dispersion of *Pareto weights*. Taking first differences therefore yields a zero on the right–hand side, since *Pareto weights* are fixed under perfect risk sharing. Beyond the cross–sectional variance implied by eq. (3), one can then also examine the variance of eq. (4), which gives:

$$Var_q(\Delta \ln c_t^{i,q}) = 0 \qquad \forall i, t. \tag{6}$$

These equations can again be used to test perfect risk sharing and to quantify deviations from it. This approach is particularly useful when multiple risk-sharing groups are observed (in our setting, multiple villages are included in the survey); and when reliable measures of income or idiosyncratic shocks are unavailable (also in our case, as noted in the introduction) making it challenging to apply the Townsend (1994) approach.

2.1 Deviations from Perfect Risk Sharing

Measurements of deviations from perfect risk-sharing in this framework can also be linked to specific imperfections that prevent complete insurance. For instance, Attanasio and Pavoni

(2011) construct a model relating imperfect information to certain departures from perfect risk sharing. More generally, in any theoretical model where an imperfection limits full insurance, observable markers of that imperfection can be related to the size of deviations from perfect risk sharing.

A natural setting for our analysis is the class of models with imperfectly enforceable contracts, such as those studied by Ligon, Thomas, and Worrall (2002). In these models, a constrained-efficient social planner maximizes participants' utility subject to both resource constraints and participation constraints, ensuring that each individual prefers remaining in the risk-sharing arrangement over deviating. Deviations typically entail future exclusion from the arrangement, along with any additional sanctions that the group can credibly enforce. When such sanctions are extremely severe, the arrangement approaches perfect risk sharing.

Under imperfect enforceability, the social planner effectively adjusts the Pareto weights in eq. (1) to satisfy participation constraints. Unlike the full risk-sharing model, where Pareto weights are exogenous, these weights now partially endogenize the dynamics of consumption allocations. When a participant is tempted to leave, their weight is increased to keep them in the arrangement.

Ligon et al. (2002) show that, in a discrete-state environment, equilibrium can be characterized by intervals corresponding to the states of the world. If the ratio of a consumer's marginal utility to the planner's Pareto weight falls within the interval for that state, transfers maintain it at the same level; if outside, transfers move it to the interval boundary, where the participant is just willing to remain. When enforceability constraints are nonbinding, the distribution of marginal utilities remains constant as under perfect risk sharing; when binding, the intervals narrow, and marginal utility distributions become more volatile.

This characterization has two key implications. First, the size of the intervals determines the amount of risk sharing achievable in equilibrium. Larger intervals imply smaller and less frequent movements in marginal utilities, which relates to the variance measure in eq. (6).

Indeed, Ligon et al. (2002) show that if the intersection of all the intervals corresponding to different states of the worlds is non-empty, the system converges to full-risk sharing. Second, stronger enforcement (larger sanctions) expands the intervals, allowing more effective risk sharing—a mechanism relevant in our context.

In our analysis, well-developed SHGs, which have accumulated significant funds and can provide effective insurance, serve as effective enforcement mechanisms. Exclusions from such groups constitute a more severe punishment on noncompliers to the established risk sharing agreements, therefore facilitating resource allocations closer to full-risk sharing. We test whether villages where the SHG program is introduced early and functions well exhibit smaller variances in consumption growth (or village-level variances of the differences in log consumption before and after the program is introduced), indicating movements towards full risk-sharing; and whether such improvements in risk-sharing are larger where program commitment, and thus contract enforceability, is stronger.

3 The Bihar Rural Livelihoods Project or *Jeevika*

We study the roll-out of one of India's largest state self-help group (SHG) programs, the Bihar Rural Livelihoods Project or *Jeevika*, between 2012 and 2014. Bihar is one of India's poorest states, and has relied extensively on *Jeevika* for poverty reduction since its inception in 2006.[4] In 2011–12 (the start of our study period), Bihar's per-capita net state domestic product was ₹13,149 (vs. a net domestic product of ₹38,048 nationally, in 2004-05 constant prices) and 34.1% of its rural population lived below the poverty line (25.7% nationally) (Government of Bihar, 2015; Government of India, 2013). This setting thus provides an opportunity to study risk-sharing where demand for informal support is likely high.

[4]In 2011, *Jeevika* was incorporated into the National Rural Livelihoods Mission (NRLM).

3.1 *Jeevika*'s Structure and Reach

Jeevika is a women's SHG program that mobilizes primarily disadvantaged women into SHGs of 10–15 members, expanding access to finance and providing basic training in literacy, numeracy, empowerment, collective action, and livelihoods.[5] SHGs meet weekly and require regular savings: in our 2019 survey of 1,217 SHGs, prescribed monthly savings averaged ₹40 per member. Savings, government funds, and bank loans form its lending pool. Each SHG receives a ₹15,000 revolving fund and a one-time ₹30,000–50,000 Community Investment Fund (Kochar et al., 2022; Hoffmann et al., 2021).

SHGs federate into Village Organizations (VOs) and VOs into Cluster Level Federations (CLFs), with CLFs typically spanning around 20 villages, with a population of about 62,000, as per the 2011 Census. A block, the administrative tier above CLFs, contains three to five clusters. After three months of regular savings, SHGs can borrow up to ₹50,000 from their VO at an interest rate of 1% per month; members can borrow from SHGs at 2% per month, far below prevailing informal rates (5% per month, on average). Borrowers are individually liable for SHG loans and collectively liable for their SHG's VO loan.

As the program rolled out, mobilization proceeded hamlet-by-hamlet, prioritizing areas with high concentrations of disadvantaged households (i.e., Scheduled Caste (*Dalit*) or Scheduled Tribe (*Adivasi*) households). By 2021, *Jeevika* included over 12 million members—68% of rural households, far above the national SHG coverage of 47%, based on 80 million SHG members nationwide and 2011 Census data. This high coverage suggests that *Jeevika* could meaningfully shape village-level risk-sharing dynamics.

3.2 *Jeevika*'s Phases and Rollout

Assignment to phases. *Jeevika* rolled out in a phased manner. Phase 1 began in 2006 in six 'high-poverty' districts. 42 blocks were selected using socio-economic indicators of

[5] Livelihood components were introduced later and are not part of the study period (Hoffmann et al., 2021).

structural disadvantage from the 2001 Census (a high share of *Dalit*/ *Adivasi* households, skewed sex ratios, low female literacy, and limited basic infrastructure) (Government of Bihar, 2011; Government of Bihar et al., 2010). Phase 2 began in 2011, expanding to 60 additional blocks, including 11 in the flood-affected Kosi region (as part of the Bihar Kosi Flood Recovery Project) (Government of Bihar et al., 2010).

Early vs. late entry villages within blocks. Within blocks, rollout prioritized more accessible villages with stronger pre-existing social capital. Our Phase 1 sample contains only late-entry villages (typically more remote, or with weaker social capital); Phase 2 sample villages were among the earliest entrants *within that phase*, with better infrastructure (and stronger social capital).

Characteristics across program phases. Phase 1 blocks are socioeconomically weaker overall, while Phase 2 (Kosi) blocks were affected by major flooding in 2008. In 2011–12, Phase 2 districts had lower per capita Gross District Domestic Product (₹9,766 vs. ₹12,275) (Government of Bihar, 2016). Baseline data show that Phase 1 households had lower per capita consumption (₹674 vs. ₹802), higher savings rates (47% vs. 37%), lower male migration (37% vs. 44%), larger villages, and lower female literacy than Phase 2 households (Tables B3 and B4). These differences imply that phase differences in initial program scale coincide with broader regional disparities. To address this, our empirical strategy relies on direct measures of initial program scale, as described in Section 4.2.

Cadre-led capacity building. A cadre of community mobilizers, CRPs or '*Jeevika didis*', were drawn from established SHGs and trained to mobilize new members, while strengthening existing groups. CRPs were deployed mainly within their own blocks, generating knowledge and capacity spillovers from early- to late-entry villages. This sequential, experience-based model relaxed capacity constraints and embedded capacity-building into the program's structure, tying SHG formation quality to pre-existing program scale (World Bank Group, 2017; Kochar et al., 2020).

SHG quality across phases. Focusing on SHGs formed in villages in the study sample (i.e., where the program rolled out in 2012 for treated areas and in 2015 for control areas), our 2019 SHG survey shows broadly similar member composition across phases: around 12 members per group; 65% *Dalit*/*Adivasi*; similar SHG vintage on average (see Table 1). The main differences appear in group quality. Phase 1 SHGs exhibited strong savings and lending activity, and were more likely to impose penalties for loan defaults. Phase 1 SHGs were also more likely to maintain records, but were less likely to hold regular weekly meetings. Because groups across phases were formed at similar times, these differences reflect lower institutional capacity and broader regional differences, rather than vintage effects. Thus, Phase 2 SHGs appear weaker in financial activity and enforcement capacity, which we argue is likely to mitigate their contribution to risk sharing. Because these patterns are cross-sectional, they should be viewed as merely suggestive.

4 Experimental Design and Variation in Institutional Capacity

In 16 blocks where the program had not yet completed coverage, *Jeevika*'s roll-out was implemented as a cluster randomized controlled trial (RCT) (Hoffmann et al., 2021). In this section, we describe the RCT and explain how it also generated plausibly exogenous variation in institutional capacity that we leverage in our analysis. The RCT spanned blocks in different program phases, which differed in pre-existing program scale (our measure of institutional capacity). Of note here is that randomization strata were defined by baseline debt levels rather than geography. As a result, these strata ended up grouping villages with different levels of pre-existing program scale, generating variation in institutional capacity available to the newly treated villages. We first describe the experimental design, then

Table 1: SHG characteristics across the two program phases in 2019

	Means			Difference
	Obs	Phase 1 Blocks	Phase 2 Blocks	in Means
SHG Characteristics				
No. of members	1209	11.651	11.806	-0.156 (0.134)
Age of SHGs (in months)	1209	51.015	52.646	-1.631 (4.195)
Dalit/Adivasi members	1209	65.539%	65.657%	-0.117 (3.232)
Agricultural HHs	1209	17.141%	15.611%	1.530 (3.304)
Members' education (years)	1209	1.775	1.052	0.723*** (0.105)
SHG Functioning				
SHG Panchsutra Score (5-point scale: meeting, saving, lending, repayment, book-keeping)	1209	2.472	1.827	0.645** (0.254)
Met at least 4 times a month last year	1209	35.678%	48.829%	-13.150* (7.096)
Saved at least ₹480 a month last year	1209	61.055%	54.871%	6.185 (7.373)
At least one loan taken by members in the last year	1209	78.894%	54.501%	24.394*** (7.133)
Has penalties for default	1209	52.764%	17.879%	34.885*** (6.352)
Maintains and has updated records	1209	18.844%	6.658%	12.186* (6.085)
SHG Finances				
Total Loanable Funds (from savings & grants) per member	1209	₹6249.094	₹4804.665	1444.429* (764.119)
Mean saving per member last year	1209	₹462.326	₹407.359	54.967* (31.041)
Mean borrowing per member last year	1209	₹4421.056	₹1415.179	3005.877*** (1014.347)
No. of SHGs	1209	398	811	
Treated (or early entry) SHGs	594	200	394	
Control (or late entry) SHGs	615	198	417	

Standard errors clustered at the program cluster level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Data from a random sample of SHGs formed prior to 2019 in 246 study villages surveyed in 2019 explain how variation in institutional capacity arises and how we measure it.

4.1 The *Jeevika* RCT

As part of the RCT, sampled Gram Panchayats (GPs) were assigned to treatment (2012 roll-out) or control (late-2014/2015 roll-out) groups. All GPs in the sample were new to the program, with the sample spanning 5 Phase 1 blocks and 11 Phase 2 blocks. Figure A1 describes the how the roll-out and its randomization were executed in Phase 1 and Phase 2 blocks. Sampled GPs in Phase 1 blocks were predominantly 'late-entry', while sampled GPs in Phase 2 blocks were 'early-entry'.

Sampling and randomization. Hoffmann et al. (2021) randomly selected 179 GPs in the study blocks.[6] One or two villages per GP were selected for data collection with probability proportional to size. From the 333 sampled villages, 8,988 households were selected using stratified random sampling (*Dalit* or *Adivasi*, and other (more privileged) caste groups). Sampling mimicked *Jeevika*'s strategy for recruitment into the program (Hoffmann et al., 2021): households were first selected from *tolas* (hamlets) where the majority of households were *Dalit* or *Adivasi*, and then from other *tolas*.

After a 2011 baseline survey, GPs were randomized into the treatment and control groups *within each block,* stratified by block and baseline high-cost ($\geq$4 % per month) debt. Sampling weights correct for the oversampling of marginalized caste households (70% of the sample). SHG formation in treated GPs occurred between January and April 2012; the endline survey was fielded between July and September 2014, after which control GPs received access to the program.

Data. In addition to household and village data from 2011 and 2014 collected by Hoffmann et al. (2021), we use data from two new sources: (1) a 2019 survey of 4-5 randomly selected

[6] One of the initial 180 was dropped due to security concerns.

SHGs in 246 of the 333 study villages; (2) NRLM's administrative MIS data. The former allows us to measure SHG functioning and quality, and the latter provides data on all SHGs and community cadre members (formation year, membership, location, cadre entry year, cluster assignment), merged with 2011 Census data. Details are in Appendix E.

Randomization balance. Tables B1 and B2 in the Appendix report randomization balance in the study sample. Column (4) shows normalized differences between the treatment and control groups, with randomization inference p-values (Fisher, 1935; Rosenbaum, 2002).[7] No variable exhibits imbalance above 0.25, noted by Imbens and Wooldridge (2009) as the threshold above which imbalance affects specification sensitivity.

Pre-existing overall evaluation results. To summarize results from Hoffmann et al. (2021), the program increased SHG participation, SHG lending and debt, and female labor force participation, with no significant effects on consumption expenditures and modest increases in durable-goods ownership. It reduced informal interest rates, the number of informal lenders, and the volume of informal loans (see Figure A4). Earlier studies focusing on phase 1 alone also found evidence of reduced household debt burdens and modest gains in women's empowerment (Datta, 2015).

4.2 Experimental Variation in Institutional Capacity

As noted in the introduction, we exploit plausibly exogenous variation in institutional capacity—measured by pre-existing program scale—to study its effect on risk sharing. This section explains where that variation comes from and how we measure it. We first document differences in institutional capacity across program phases, then show how the randomization generates variation in capacity at the cluster level.

[7] Randomization inference implemented following Heß (2017) in Stata. Normalized differences, $\hat{\Delta}_{ct} = \frac{\bar{X}_t - \bar{X}_c}{\sqrt{(s_t^2+s_c^2)/2}}$, provide scale-free measures of covariate imbalance (Imbens and Rubin, 2015).

Institutional capacity varied sharply across program phases. The program rolled out in two phases. Phase 1 blocks began in 2006 and had accumulated 186 SHGs per 100,000 people by 2010. Phase 2 blocks only began rolling out in 2010–11 and had almost none (0.26 per 100,000). This gap in SHG presence reflects a corresponding gap in institutional capacity. A further indicator is the stock of Community Mobilizers (CRPs), who are drawn from existing SHGs: Phase 1 blocks entered the study with far greater mobilization capacity than Phase 2 blocks (see Figure A2 and Figure A3 in the Appendix).

The randomization generates variation in institutional capacity at the 'operating area' level. As discussed in Section 4.1, the experiment was randomized within blocks, and blocks spanned both Phase 1 and Phase 2 regions. Within each block, the program was implemented within smaller geographic units or 'operating areas', to which administrative staff and community cadre members were assigned and within which they operated. In *Jeevika's* own terminology, this 'operating area' is a program 'cluster', and going forward, we refer to it as a 'cluster'. Importantly, this 'cluster' was unique to *Jeevika*, and not a geographic unit relied on for any other policy or administration. In addition, the experimental design of the RCT produced variation in institutional capacity across different clusters. We exploit this variation for identification, and explain precisely how the variation in institutional capacity arises below.

A cluster consisted of around 20 contiguous Gram Panchayats (GPs), and our sample covers 50 such clusters. Control GPs were excluded from mobilization during the study period. The key insight is that pre-existing program resources, i.e., staff and community mobilizers, had to be deployed across whichever GPs in a cluster were assigned to treatment. The effective area over which those resources had to operate therefore depended on how many GPs in a cluster were assigned to control, which was determined by the randomization.

As a result, two clusters with identical pre-existing institutional capacity could face very different mobilization burdens. If most GPs in a cluster were assigned to treatment, resources had to be spread over a large area. If many GPs were assigned to control, the same resources were concentrated over a smaller area, enabling more intensive engagement with each treated village.

Crucially, this variation in treatment-control composition across clusters was not by design. The randomization was stratified by block and by GP-level mean high-cost debt at baseline, and not by geography. Strata therefore grouped two to three GPs drawn from potentially different clusters, so the share of treated GPs within any given cluster was an outcome of the randomization rather than something fixed in advance.

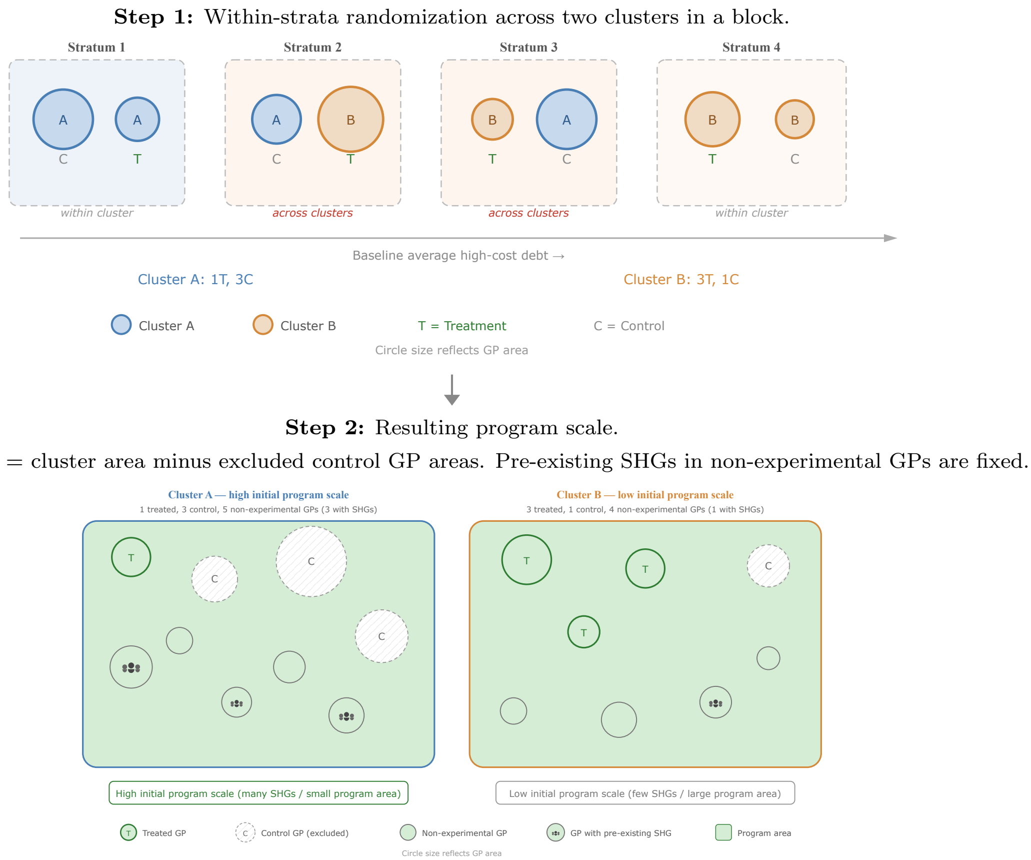
Measuring institutional capacity. We measure initial program scale as the number of SHGs established before the experiment (in 2010) relative to the target geographic area over which mobilization was to take place, excluding control GPs:

$$Scale_c = \frac{\text{No. of pre-existing SHGs}_{2010,c}}{\text{Target geographic area for mobilization}_c} \tag{7}$$

A higher value indicates that program staff and community resource persons could engage more intensively with each treated village—visiting more frequently, following up more consistently, and building a stronger institutional presence.

The numerator, SHG presence in 2010, is pre-determined. The identifying variation in this measure comes from the denominator—the target geographic area (in hectares) over which mobilization was to occur, i.e., the cluster area minus the area of control GPs, which were excluded from program activities. Figure 1 illustrates this directly: in Step 2, Cluster A has many pre-existing SHGs concentrated over a small program area (high scale), while Cluster B has few SHGs spread across a large program area (low scale). Since treatment assignment was random, the size of the excluded control area, and therefore the denominator, varies randomly across clusters. The strata fixed effects in our empirical specifications, which group GPs within the same block and with similar baseline mean high-cost debt, absorb

Figure 1: Experimental variation in initial program scale across clusters



Note: Experimental GPs in a block are grouped into strata by baseline average high-cost debt. Strata can cross cluster boundaries (Step 1), so the treatment–control split within each cluster is random. As a result, the number of control GPs in a cluster is random, and the size of the control GPs is random. The program operates over the cluster area excluding control GPs; pre-existing SHGs are fixed. $Initial\ program\ scale = \frac{pre-existing\ SHGs}{cluster\ area\ -\ control\ GP\ area}$.

any remaining block-level confounds and ensure that comparisons are made across clusters facing similar baseline conditions. Results are robust to alternative scale definitions and the inclusion of 2001 census controls.[8]

5 Does *Jeevika* Improve Risk Sharing?

In this section, we present our main empirical results. In Section 5.1, we start by presenting our test of perfect risk sharing, looking at changes in the within-village cross sectional vari-

[8]The results reported in the Appendix are obtained using alternative definitions, based on per-woman population or per-target member normalizations. Some definitions also use pre-experiment numbers of CRPs instead of SHGs.

ance of consumption. In Section 5.2, we present evidence that the *Jeevika* intervention did increase risk sharing, at least in Phase 1 villages. In Section 5.3, we analyze possible mechanisms that could explain deviations from perfect risk sharing, focusing on the institutional capacity of the SHG program. The mediation analysis we perform in this section will lead us, in Section 6, to consider the link between institutional capacity and the quality of SHGs, and, implicitly, their value as a risk sharing mechanism.

5.1 Is There Perfect Risk Sharing?

Following the framework laid out in Section 2, we examine how the program affected the *within-village* variance of household consumption growth between 2011 and 2014, which provides a direct measure of informal village-level risk sharing. We interpret log consumption as an empirical proxy for (log) marginal utility, and we label the village-level variance of consumption growth between 2011 and 2014 for all sample households, i, in a village, v, in a Gram Panchayat, g, in cluster, c, as $Var_{vgc}(\triangle_{2011}^{2014} \log c^{ivgc})$. Our conceptual framework indicates that in the perfect risk-sharing case, $Var_{vgc}(\triangle_{2011}^{2014} \log c^{ivgc}) = 0$.

Table 2 contains a first test of perfect risk sharing and shows, as one might expect, that rural Bihar is not a perfect risk-sharing world. The rejection of the perfect risk sharing hypothesis is indicated by the fact that the variance of the changes in individual marginal utilities, as approximated by consumption growth is, on average, significantly different from zero. This result holds when we consider all villages together but also when we separate Phase 1 and 2.

A word of caution should be added to an interpretation of these results as rejection of perfect risk sharing. There are two possibilities that would give rise to the non-zero variability of consumption growth we see in Table 2 even in a world with perfect risk sharing. First, in the presence of measurement error with time varying variance, one could observe significant changes in the cross sectional variance of individual consumption growth even when the variance of actual log consumption is constant. Second, if log consumption does not constitute a good approximation of marginal utility, which could also be affected by other variables, such as observed and unobserved taste shocks, one could still observe changes to the cross sectional variance of log consumption even when the market for full insurance is perfect. While these are plausible objections, tests of whether the program changes these measures would still be valid, unless one would think that measurement error or taste shocks are themselves affected by the *Jeevika* program.

Table 2: Is the variance of consumption growth zero (in control group villages)?

Variable	Phases	Mean	t-stat	p-value	N
All Households	1 & 2	0.243	27.825	0.000	169
	1	0.226	15.932	0.000	58
	2	0.252	22.934	0.000	111

T-tests of variance of consumption growth means against zero in control group villages.

5.2 Does Access to *Jeevika* Improve Risk Sharing?

The next step in our analysis is testing the hypothesis that the SHG intervention affects the amount of risk sharing in a village. We interpret larger levels of the cross sectional variance of consumption growth as larger deviations from perfect risk-sharing. To test whether these deviations have been affected by the program we compare this measure in treatment and control GPs.

Empirical Strategy. In particular, we estimate:

$$Var_{vgc}(\triangle_{2011}^{2014} \log cons^{ivgc}) = \beta_1 \mathbf{T}_{gc} + \beta_2(\mathbf{T}_{gc} \times \textbf{Phase 1}_c) + \mathbf{X}_{vgc}^{'}\beta_6 + \mathbf{S} + \epsilon_c \quad (8)$$

where $\triangle_{2011}^{2014} \log cons^{ivgc}$ is the change in log consumption (or consumption growth) for household i in village v, GP g, and program cluster c; $Var_{vgc}(\triangle_{2011}^{2014} \log cons^{ivgc})$ is the village-level variance of consumption growth across all households in the village; $\mathbf{T}_{gc}$ is an indicator for a GP's random assignment to treatment; $\textbf{Phase 1}_c$ is an indicator for whether the village is in a phase 1 block; and $\mathbf{S}$ denotes strata fixed effects. Some specifications include $\mathbf{X}_{vgc}$, a vector of pre-treatment village-level measures from the 2001 census used by the government to assign regions to program phases (share of *Dalit*/*Adivasi* households, sex ratios, female literacy rate, infrastructure index). Robust standard errors are clustered at the program-cluster level.

Table 3: Impact of access to *Jeevika* on village risk sharing

	Variance of change in log consumption			
	(1)	(2)	(3)	(4)
Treatment	-0.018	-0.019	-0.003	-0.003
	(0.012)	(0.011)	(0.016)	(0.016)
Treatment × Phase 1			-0.044**	-0.049**
			(0.019)	(0.021)
Controls	No	Yes	No	Yes
Obs	333	333	333	333
Clusters	50	50	50	50
Mean (Omitted Group)	0.243	0.243	0.252	0.252

Notes: The outcome is the variance of change in log consumption, and is computed relying on sampling weights to reconstitute caste composition of the village. Phase 1 is an indicator which takes the value 1 when a village is in the program's first phase. All specifications include randomization strata fixed effects. Regressions in even columns include controls (from the 2001 census) for the village Dalit and Adivasi population share, village female literacy rate, village sex ratio, an index of village infrastructure. Robust standard errors clustered at the program cluster level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Results. The estimates in Table 3 point to two interesting features. First, when considering the entire sample of GPs, as done in columns 1 and 2, we find that, while a negative point estimate of the treatment coefficient indicates a reduction in our measure of deviation from perfect risk sharing, these coefficients are not significantly different from zero. Second, when we let the impact of the program to be different in Phase 1 and Phase 2 GPs, we notice that Phase 1 GPs experience a significant reduction in the variance of individual consumption growth. Such coefficients represent a 20% reduction in our measure of deviations from perfect risk sharing. As we have argued, in Phase 1 GPs, the program had much more developed institutional capacity. It is to this that we now turn and asks whether the improvement in risk sharing comes from an increased institutional capacity.

5.3 Does *Jeevika*'s Institutional Capacity Improve Risk Sharing?

The results in Table 3 show that in Phase 1 areas, *Jeevika* does improve risk sharing. As shown in Table 1, the program seems to be better developed and established in Phase 1 areas. Our next step, therefore, consists in relating the institutional capacity and the scale of the program to the improvement in risk sharing.

Empirical strategy. In this section, we test whether better institutional capacity, represented by larger initial program scale, leads to greater improvements in risk sharing among treated villages. As we discussed in Section 4.2, the variability in scale is induced by the way the RCT developed and can be considered as exogenous.[9] Therefore, we estimate:

$$
\begin{aligned}
Var_{vgc}(\triangle_{2011}^{2014} \log cons^{ivgc}) = \; & \beta_1 \mathbf{T}_{gc} + \beta_2(\mathbf{T}_{gc} \times \mathbf{Phase\ 1}_c) + \beta_3 \mathbf{Scale}_c + \\
& + \beta_4(\mathbf{T}_{gc} \times \mathbf{Scale}_c) + \beta_5 \mathbf{Scale}_c^2 + \beta_6(\mathbf{T}_{gc} \times \mathbf{Scale}_c^2) + \mathbf{X}_{vgc}^{'} \beta_7 + \mathbf{S} + \epsilon_c \qquad (9)
\end{aligned}
$$

where $\triangle_{2011}^{2014} \log cons^{ivgc}$ is the change in log consumption (or consumption growth) for household i in village v, GP g, and program cluster c; $Var_{vgc}(\triangle_{2011}^{2014} \log cons^{ivgc})$ is the village-level variance of consumption growth across all households in the village; $\mathbf{T}_{gc}$ is an indicator for a GP's random assignment to treatment; $\mathbf{Phase\ 1}_c$ is an indicator for whether the village is in a phase 1 block in certain specifications; $\mathbf{Scale}_c$ is the measure of initial program scale defined in eq. (7); and $\mathbf{S}$ denotes strata fixed effects. Some specifications include quadratic terms in scale and its interaction with treatment, and some specifications include $\mathbf{X}_{vgc}$, a vector of pre-treatment village-level measures from the 2001 census used by the government to assign regions to program phases (share of *Dalit*/*Adivasi* households, sex

[9] Scale variation, however, is not *completely* explained by the RCT, otherwise, eq. (9) would not be identified. One could argue that the treatment had different *intensity*, which is articulated through the exogenously varying scale.

ratios, female literacy rate, infrastructure index). Robust standard errors are clustered at the program-cluster level.

Because only treated villages had access to the program during this period, our parameters of interest are β_3 to β_6, capturing the interaction between treatment and initial program scale. At the same time, we consider how the estimate of the direct treatment effects (β_1 and β_2) change when we consider the effect of scale.

Results. Table 4 presents our empirical results on the relation between the impact of the program on risk sharing and the scale of the program. Recall from Section 2 that perfect risk sharing implies zero variance of consumption growth; a negative coefficient therefore indicates an improvement in risk sharing.

In Table 4, the coefficient on the interaction between treatment and initial program scale is negative and significant across all specifications, indicating that the program's effect on risk sharing increases with institutional capacity. The estimates for the specification in column 2, which presents a linear model with several controls, imply that an additional pre-existing SHG per 100 hectares to be mobilized reduces the variance of consumption growth by 7% in treated villages, a meaningful improvement in risk sharing.[10]

This finding is robust to definitions of initial program scale alternative to what defined in eq. (7), as we document in Table B7 in the Appendix.[11] The interaction term remains negative and significant across these specifications.

Reassuringly, the stability of the interaction coefficient across odd and even columns in Table 4 (without and with the four 2001 census controls used to assign areas to early roll-

[10]In Table B6 in the Appendix, we consider the cross-sectional variance of consumption growth (or changes in log consumption) within caste categories in a village, i.e., within Dalit/Adivasi households and non-Dalit/Adivasi households within a village. We find similar results when we consider caste categories within a village as the relevant risk sharing groups.

[11]We measure institutional capacity using either pre-existing SHGs or pre-existing CRPs, and define the mobilization target as geographic area, number of women to be mobilized to reach program targets, or number of women in eligible areas based on the 2011 census.

out) further suggests that observables correlated with early program assignment do not drive the estimates, while any level differences in consumption variance across higher- and lower-scale areas are absorbed by the main effects of scale, making it unlikely that fixed household risk preferences drive the interaction with randomly assigned treatment.

Columns (3) and (4) of Table 4 and the even columns of Table B7 include quadratic terms to examine whether the relationship is non-linear. The quadratic specification reveals diminishing returns, with the marginal effect of initial program scale on risk sharing declining at higher levels of scale.

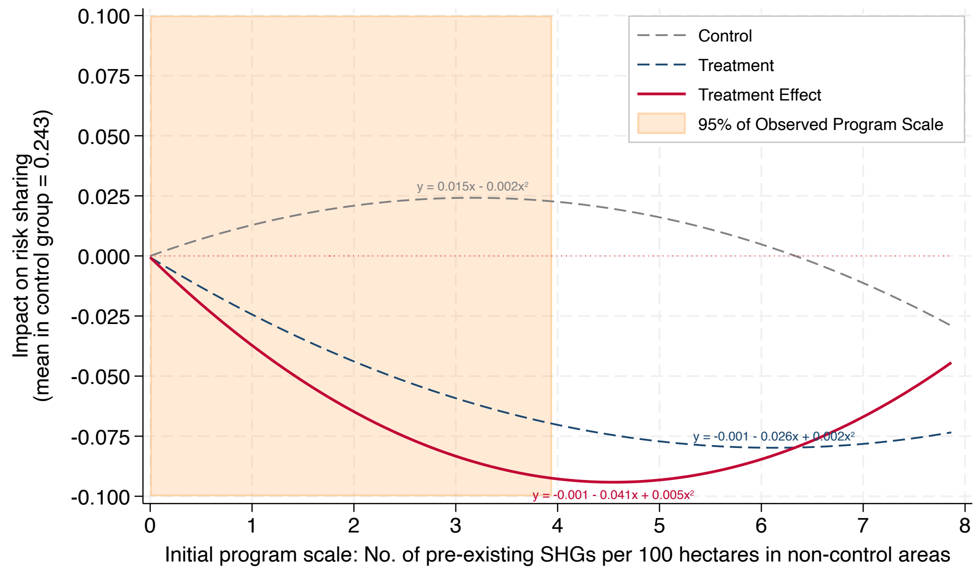
Table 4: Impact of *Jeevika*'s institutional capacity on village risk sharing

	Variance of change in log consumption			
	(1)	(2)	(3)	(4)
Treatment	-0.007	-0.006	-0.001	-0.001
	(0.014)	(0.014)	(0.015)	(0.015)
Initial program scale	-0.000	-0.000	0.014	0.015
	(0.003)	(0.004)	(0.009)	(0.013)
Treatment × Initial program scale	-0.014**	-0.017***	-0.040***	-0.041**
	(0.006)	(0.006)	(0.014)	(0.015)
Initial program scale squared			-0.002**	-0.002
			(0.001)	(0.002)
Treatment × Initial program scale squared			0.005***	0.005**
			(0.002)	(0.002)
Controls	No	Yes	No	Yes
Obs	333	333	333	333
Clusters	50	50	50	50
Mean (Control Group)	0.243	0.243	0.243	0.243

Notes: The outcome is the variance of change in log consumption, and is computed relying on sampling weights to reconstitute caste composition of the village. Initial program scale is defined in Section 4.2. All specifications include randomization strata fixed effects. Regressions in even columns include controls (from the 2001 census) for the village Dalit and Adivasi population share, village female literacy rate, village sex ratio, an index of village infrastructure. Robust standard errors clustered at the program cluster level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 2 plots the estimated treatment effect across the range of initial program scale observed in the data. The treatment effect is negative (indicating improved risk sharing) across the range of initial program scale. The effect is larger with larger scale throughout 95% of observed data, though the marginal improvement diminishes at higher scale levels. This pattern is consistent with capacity constraints: at very high scale, administrative resources may be stretched across too many existing groups, reducing the intensity of support for newly formed SHGs. The linear interaction in columns (1) and (2) remains our preferred specification.

Figure 2: Non-linear effect of initial program scale on risk sharing by treatment status



Finally, Table 5 presents a mediation analysis to establish whether difference in institutional capacity and scale can explain the difference in impact between Phase1 and Phase 2. Columns (1)–(2) show, again, that treatment effects differ significantly between Phase 1 and Phase 2 blocks. However, once we control for initial program scale and its square, the phase difference coefficient changes sign and becomes insignificant (columns 5–6), indicating that variation in institutional capacity, rather than other regional differences, explains the heterogeneity in treatment effects across phases when we account for possible diminishing returns due to capacity constraints.

As strata fixed effects absorb all cross-phase variation, Table 5 shows that initial program scale, rather than phase-level differences, drives the phase heterogeneity. This evidence addresses the concern that the institutional capacity results are driven by the fact that clusters in Phase 1 blocks mechanically have higher institutional capacity than those in Phase 2 blocks.

Table 5: Impact of *Jeevika*'s insitutional capacity on village risk sharing

	Variance of change in log consumption					
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Treatment × Phase 1	-0.044**	-0.049**	-0.035	-0.022	0.059	0.085*
	(0.019)	(0.021)	(0.027)	(0.031)	(0.043)	(0.047)
Initial program scale			-0.004	-0.002	0.028**	0.036**
			(0.003)	(0.004)	(0.013)	(0.014)
Treatment × Initial program scale			-0.005	-0.011	-0.076**	-0.092***
			(0.007)	(0.008)	(0.029)	(0.030)
Initial program scale squared					-0.004**	-0.005***
					(0.001)	(0.002)
Treatment × Initial program scale squared					0.009***	0.010***
					(0.003)	(0.003)
Controls	No	Yes	No	Yes	No	Yes
Obs	333	333	333	333	333	333
Clusters	50	50	50	50	50	50
Mean (Control Group, Phase 2)	0.252	0.252	0.252	0.252	0.252	0.252

Notes: The outcome is the variance of change in log consumption, and is computed relying on sampling weights to reconstitute caste composition of the village. Initial program scale is defined in Section 4.2. Regressions in even columns include controls (from the 2001 census) for the village Dalit and Adivasi population share, village female literacy rate, village sex ratio, an index of village infrastructure. Robust standard errors clustered at the program cluster level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We note that since our measure of institutional capacity is plausibly exogenous, the mediation analysis indicates that the phase differences are driven by institutional capacity differences. If both the phase differences and institutional capacity differences were significant, it would indicate that there might be geographical differences across phases, over and above which, there also emerge institutional capacity differences.

6 How Does *Jeevika* Improve Risk Sharing?

In this section, we discuss some empirical evidence on the channels through which *Jeevika* may affect risk sharing. The theoretical framework we sketched in Section 2 suggests that the program's institutional capacity might improve risk sharing through informal transfers because it increases the quality of SHGs, and therefore their value to their members. If the SHG members share risk, this increased value can facilitate more risk sharing as it offers a more effective enforcement mechanism.

6.1 Risk Sharing and *Jeevika*'s Other Impacts

Table 6 provides the most direct evidence that the impact of *Jeevika* on the cross sectional variance of log consumption growth (and a reduction in its movements) reflects improvements in risk sharing. In particular, we consider potential mediators of the impact on the cross sectional variance of log utility that we have documented. Columns (1) and (2) show that access to *Jeevika* increases the likelihood that households approach their social networks (individuals outside of immediate relatives) during food and health emergencies by 5.820 and 7.795 percentage points. This effect does not vary with initial program scale, indicating that the program builds networks across the board. The scale-dependent improvement in risk sharing therefore operates over and above this, rather than only through it.

However, we note that emergency borrowing declines significantly with initial program scale among treated households (columns (3) and (4)). If scale simply expanded credit availability, we would expect more emergency borrowing in higher-scale areas, not less. Since columns (3) and (4) condition on households that actually experienced the relevant shock, the decline reflects reduced reliance on borrowing among the affected, rather than reduced access to it. This suggests that better-functioning SHGs provide households with sufficient insurance—possibly through mutual support and access to group savings, apart from lending.

The available evidence, which we report in the Appendix, shows that SHGs provide members

Table 6: Impact of *Jeevika*'s institutional capacity on networks & insuring shocks

	Do you approach network for		Did you borrow for	
	Food Shortage (%)	Health Shock (%)	Food Shortage (%)	Health Shock (%)
	(1)	(2)	(3)	(4)
Treatment	5.820**	7.795***	2.331	1.704
	(2.591)	(2.809)	(2.193)	(1.807)
Initial program scale	-0.641	-1.143	0.442	2.007***
	(0.736)	(0.930)	(0.508)	(0.429)
Treatment × Initial program scale	-0.714	0.311	-3.350***	-1.449*
	(1.381)	(1.185)	(0.920)	(0.841)
Obs	8988	8988	8987	8987
Clusters	50	50	50	50
Mean (Control group)	49.886	45.046	48.913	71.900

Notes: Outcomes in columns (1)–(2): whether the respondent approaches members outside immediate relatives or moneylenders for help with food shortages or health shocks. Outcomes in columns (3)–(4): whether the household borrowed to buy food or meet health expenses in the preceding 12 months, conditional on experiencing the relevant shock. All outcomes in percentage points. Initial program scale is defined in equation 7 in Section 4.2. All regressions include strata fixed effects, household-level baseline controls from Hoffmann et al. (2021), and village-level controls from the 2001 census (Dalit/Adivasi population share, female literacy rate, sex ratio, infrastructure index). Columns (3)–(4) additionally control for incidence of the relevant shock. Robust standard errors clustered at the program cluster level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

with access to savings, credit, and strengthen networks. Table C1, for instance, shows that access to *Jeevika* substantially increases SHG borrowing, reduces informal borrowing, and lowers interest rates, while Table B8 (column 1) shows that it increases the incidence of savings. These are channels through which group membership may improve risk sharing, not competing explanations. If improved credit or savings access through SHGs reduces the variance of consumption growth, that is a mechanism through which risk sharing operates, not an alternative to it. However, none of these effects vary significantly with initial program scale among treated households. Credit and savings access therefore cannot fully explain why higher-scale areas exhibit greater improvements in risk sharing. We argue below that the additional leverage of higher-scale programs operates instead through the enforcement of informal insurance arrangements.

Furthermore, an income channel is unlikely to drive these results since *Jeevika*'s livelihoods activities had not commenced during the study period (2011–2014), and the interaction between treatment and initial program scale does not significantly affect mean consumption,

consumption growth, or consumption assets (Table B8). The interaction between treatment status and initial program scale also does not significantly impact women's labor force participation (Table C4), making income unlikely to be the main channel driving the results.

The theoretical framework in Section 2 implies that in a situation where insurance contracts are not easily enforced, risk sharing can improve when the value of continuing group membership increases, as possible exclusion could be an effective enforcement mechanism. Higher initial program scale may achieve this by enabling better-functioning SHGs, exclusions from which could be very costly. This enforcement channel, rather than expanded credit, savings, or income, is the most plausible explanation for why improvements in risk sharing are concentrated in treated areas with higher initial program scale Section 5.3. We now investigate potential mechanisms behind this result.

6.2 Does *Jeevika*'s Institutional Capacity Improve SHG Quality?

In Section 5.3, we have shown that the effect of the program on risk sharing increases with institutional capacity—inducing cross-sectional consumption allocations that are closer to allocations consistent with perfect risk sharing within villages where initial program scale is higher. Our proposed explanation centers on the participation constraint in models of informal insurance with limited commitment: risk sharing improves when the value of continued membership increases relative to the outside option. In this case, the group we consider is a village, but more specifically it is households in the village who participate in the SHG program (in our study sample, around 55% of village households participate in the program, and in 2021, 68% of rural households in Bihar were part of *Jeevika*).

Higher initial program scale can raise the value of continued group membership if SHGs in higher-capacity areas function better, both financially and in terms of governance. First, better-functioning SHGs accumulate larger savings pools, increasing loanable funds available to members. This raises the explicit value of continued group membership and, correspondingly, the cost of exclusion—members who deviate from informal agreements risk losing access to accumulated savings.

Second, better functioning groups likely maintain regular meetings, consistent bookkeeping, and reliable saving, lending and repayment. These features reflect members' incentives to sustain cooperation, and they signal continued access to group resources: good group governance ensures that funds continue to flow into and back into the groups as loans are repaid and savings accumulate. Moreover, the government relies on these specific five features—the *panchsutra*—to score SHG quality, and scores determine access to certain future grants and bank loans. Higher scores thus translate directly into greater future resources, further raising the value of group membership in well-functioning groups.

Both channels tighten the participation constraint in risk sharing contracts with imperfect enforceability by increasing what members stand to lose from exclusion, enabling greater risk sharing. This mechanism builds on the framework in Section 2, where the totality of SHG savings across the risk-sharing group strengthens enforcement of informal agreements.

Empirical strategy. To verify that what we propose is a plausible narrative, we test whether higher initial program scale leads to better functioning SHGs using data from our 2019 SHG survey. The sample on which we perform this exercise includes 509 SHGs that were formed before the end of the experiment (October, 2014)—82.5% of sampled SHGs in treated GPs, and 2.4% of sampled SHGs in control GPs. To relate SHGs quality to its scale we estimate:

$$Y_{svgc} = \beta_1 \mathbf{Scale}_c + \mathbf{X}'_{vgc}\beta_2 + \mathbf{B} + \epsilon_c \tag{10}$$

where Y_{svgc} is an outcome measuring SHG functioning for group s in village v, GP g, and program-cluster c; $\mathbf{Scale}_c$ is the measure of initial program scale defined in eq. (7); and $\mathbf{B}$ denotes block fixed effects. In some specifications we include $\mathbf{X}_{vgc}$, a vector of pre-treatment village-level measures from the 2001 census used by the government to assign regions to program phases (share of *Dalit*/*Adivasi*) households, sex ratios, female literacy rate, infrastructure index). Robust standard errors are clustered at the program-cluster level. Since we test hypotheses on related outcomes, we account for multiple hypothesis testing by presenting adjusted p-values using the Holm (1979) method.

Table 7: Group quality in 2019 and *Jeevika*'s institutional capacity for SHGs formed in the experimental period

	Loanable Funds per Member (₹)		High Panchsutra (%)		SHG Quality Index	
	(1)	(2)	(3)	(4)	(5)	(6)
Initial program scale	611.729***	634.331***	4.889**	3.660	0.148***	0.132***
	(189.874)	(216.547)	(2.394)	(2.599)	(0.037)	(0.038)
Holm (1979) adjusted p-value	0.005	0.011	0.048	0.167	-	-
Controls	No	Yes	No	Yes	No	Yes
Obs	509	509	509	509	509	509
Clusters	39	39	39	39	39	39
Mean	7777.760	7777.760	17.485	17.485	0.371	0.371

Notes: Outcomes, described in column headers, come from a 2019 primary follow-up survey of a sample of SHGs in study villages. The results presented include 509 SHGs in the study sample that were formed during the experiment (between Jan, 2012 and October, 2014, the end of the endline survey). Initial program scale is defined in Section 4.2. Even-numbered columns include village-level controls from the 2001 census: Dalit and Adivasi population shares, female literacy rate, sex ratio, and an infrastructure index. All specifications include block fixed effects and an indicator for panchayat treatment assignment. Adjusted p-values account for multiple hypothesis testing using the (Holm, 1979) method. In columns (1) and (2), the outcome, 'loanable funds per-member' is the average funds available per member from pooled savings and grants, in rupees. In columns (3) and (4), the outcome, 'high panchsutra' is an indicator that equals one if the SHG scores at least 4 out of 5 on the panchsutra index (one point each for regular meetings, bookkeeping, lending, repayment, and saving). In columns (5) and (6), SHG Quality Index is a standardized index, averaged over the two outcomes. Robust standard errors clustered at the program cluster level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Results. Table 7 examines three measures of SHG quality. The first is loanable funds per member—total savings and grants accumulated by the SHG through December 2018, divided by the number of group members. We exclude bank loans from this measure since the data do not indicate outstanding balances (only cumulative borrowing), and because bank loan access is itself determined by panchsutra scores. The second measure is *high panchsutra*, an indicator for whether the group adheres to at least four of the five panchsutra principles: regular meetings, saving, lending, repayment, and proper record-keeping. The third is a standardized index averaging the first two outcomes to provide a composite measure of SHG quality.

Columns (1) and (2) show that higher initial program scale leads to significantly greater loanable funds per member—an additional pre-2011 SHG per 100 hectares to be mobilized increases funds by ₹634, consistent with our hypothesis that better-resourced areas produce better-functioning groups. Columns (3) and (4) suggest that higher initial program scale also increases panchsutra adherence by around 4 percentage points, though this result is not robust to the inclusion of controls or to multiple hypothesis testing adjustment. The composite index in columns (5) and (6) indicates that higher initial program scale improves overall SHG quality by 0.13–0.14 standard deviations. Notably, these outcomes are measured in 2019, 5-7 years after group formation, indicating that the effects of initial program scale persist well beyond the experimental period.

6.3 Does SHG Quality Account for the Effect of *Jeevika's* Institutional Capacity on Risk Sharing?

The results we have shown so far establish that higher initial program scale improves risk sharing (Section 5.3) and produces better-functioning SHGs (Section 6.2). While we cannot provide definitive evidence on this mechanism, we now present suggestive evidence, discussed in more detail in Appendix D.2, that SHG quality might indeed be the channel linking the two.

We use a two-stage least squares specification relating our measure of risk sharing (the village level variance of consumption growth) to village-level SHG quality instrumented with the interaction between treatment and initial program scale. The three proxies for SHG quality described above serve as alternative endogenous variables, and the second-stage outcome is the village-level variance of consumption growth. The IV estimates are consistent with the proposed mechanism: higher predicted SHG quality reduces the variance of consumption growth, with a negative coefficient in all six specifications. This effect is statistically significant at conventional levels in column (2) and marginally significant in the remaining five columns.

Table 8: Impact of village SHG quality on village risk-sharing (IV)

	Variance of change in log consumption					
	Endogenous regressor is *Loanable Funds*		Endogenous regressor is *High Panchsutra Score*		Endogenous regressor is *SHG Quality Index*	
	(1)	(2)	(3)	(4)	(5)	(6)
	Second-stage					
Loanable Funds	-0.096*	-0.113**				
	(0.050)	(0.053)				
Panchsutra Score			-0.206*	-0.251*		
			(0.120)	(0.129)		
SHG Quality Index					-0.053*	-0.068*
					(0.032)	(0.035)
	First-stage co-efficients					
Initial program scale × Treatment	0.146***	0.145***	0.068**	0.066**	0.263***	0.241***
	(0.052)	(0.054)	(0.027)	(0.027)	(0.100)	(0.092)
Controls	No	Yes	No	Yes	No	Yes
Obs	333	333	333	333	333	333
Clusters	50	50	50	50	50	50
Olea and Pflueger (2013) F-statistic	7.869	7.108	6.395	6.096	6.855	6.887
Critical Values (% worst case bias)						
τ=5%	37.418	37.418	37.418	37.418	37.418	37.418
τ=10%	23.109	23.109	23.109	23.109	23.109	23.109
τ=20%	15.062	15.062	15.062	15.062	15.062	15.062
τ=30%	12.039	12.039	12.039	12.039	12.039	12.039
Anderson-Rubin F-stat	5.218	7.825	5.218	7.825	5.218	7.825
Anderson-Rubin p-value	0.027	0.007	0.027	0.007	0.027	0.007

Notes: The **excluded instrument** is Treatment × Initial program scale. Initial program scale is defined in Section 4.2. The **outcome** is the variance of change in log consumption, and is computed relying on sampling weights to reconstitute caste composition of the village. In columns (1) and (2), the **endogenous regressor** is the *village-level accumulated per-household loanable funds with all SHGs ('00,000) at the end of 2014*. In each village, the total per-household loanable funds at the end of 2014 is constructed using the mean monthly per-member loanable funds (from member savings and grants) across all SHGs in a village from the 2019 SHG survey, the number of members in each SHG, the number of SHGs and the age of SHGs at the end of 2014 from the MIS data. In columns (3) and (4), the **endogenous regressor** is the *share of SHGs in the village with a high panchsutra score (at least 4 out of 5)*. This assumes that the panchsutra scores at the end of 2014 is the same as in 2019. In columns (5) and (6), the **endogenous regressor** is the *mean SHG quality in a village*. SHG quality is a standardized index constructed as the mean of the standardized quality variables in the 2019 SHG survey: (i) accumulated village loanable funds per household; (ii) share of SHGs with high panchsutra scores in the village. The first-stage coefficient rows report the coefficients on the excluded instrument in the first-stage regression (Treatment × Scale), where the endogenous regressor is regressed on the instrument and controls. **Controls or included instruments** are: treatment, program scale, and in even numbered columns, village-level controls from the 2001 census for Dalit and Adivasi population shares, female literacy rate, sex ratio, and an infrastructure index. Weak identification is assessed using the Olea-Flueger F-statistic, which accounts for heteroskedastic errors. (Olea and Pflueger, 2013). Robust standard errors clustered at the program cluster level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The fact that the results hold across all three proxies for SHG quality is reassuring. However, since the first stage is not strong, we interpret this as suggestive evidence that SHG quality is the channel linking initial program scale and risk sharing, consistent with the participation-constraint mechanism outlined in Section 2. The weak first stage means the magnitude of these effects is not reliably estimated; the negative sign and its statistical significance, however, are robust to weak identification, since inference based on the Anderson-Rubin statistic does not rely on a strong first stage. The reduced-form result, that higher initial program scale improves risk sharing among treated villages, remains the primary finding and is identified independently.

7 Conclusion

Access to savings and lending groups improves village-level risk sharing where institutional capacity is sufficient. A one-unit increase in institutional capacity reduces the variance of consumption growth by 7% among treated households, an effect mediated by better-functioning groups with larger accumulated funds and stronger governance. The mechanism is consistent with models of informal insurance under limited commitment: higher-quality groups raise the value of membership, relaxing the participation constraints that limit risk sharing.

These findings have direct implications for the design of programs operating at scale. *Jeevika*'s phased rollout approach, particularly seen in its phase 1 regions where rollout spanned several years, drew on community mobilizers and institutional capacity from already existing and experienced SHGs to aid the expansion process. This embedded capacity building directly into the expansion process, with meaningful consequences. In clusters with higher pre-existing program-scale, newly formed SHGs were 'higher-quality'. In addition, where initial program scale was higher, households experienced not just the mean credit market and network expanding effects observed in all areas with access to the program, but also larger improvements in informal risk-sharing. The lesson here is that the sequencing of program expansion within administrative units can be designed to generate capacity spillovers from early- to late-entry areas. Programs that expand uniformly across all areas simultaneously forgo these spillovers.

Our study addresses several concerns about external validity common to program evaluations (Muralidharan and Niehaus, 2017; Banerjee et al., 2017a). The sample spans multiple program phases and administrative blocks, and sampling mimicked Jeevika's mobilization strategy, making results broadly representative of rural Bihar. Importantly, we evaluate a program implemented at scale rather than an early pilot, spanning regions with different levels of pre-existing institutional capacity at the time of the experiment. This variation allows us to characterize how program impacts differ—risk-sharing benefits are concentrated in higher-capacity areas, informing expectations for contexts where institutional infrastructure is less developed.

However, certain limitations remain. Our measure of institutional capacity relies on Jeevika's community-based mobilization model, so findings may not generalize to programs using other implementation approaches. In states with weaker mobilization or competition from microfinance institutions, participation constraints may operate differently. Programs also mature, as *Jeevika* has with an expansion in its scope of activities. So, impacts on mean and distributional outcomes may evolve accordingly.

Even so, our core finding, that the insurance value of savings and lending groups depends on institutional quality, and that this quality can be shaped by program design, is likely to generalize beyond this setting. For policymakers, the implication is clear: sequencing program expansion to build institutional capacity before scaling can yield lasting returns.

References

Albarran, Pedro and Orazio P. Attanasio (2003). "Limited commitment and crowding out of private transfers: Evidence from a randomised experiment," *The Economic Journal*, 113: C77–C85. [2]

Ambrus, Attila, Wayne Y. Gao, and Pau Milán (2022). "Informal risk sharing with local information," *The Review of Economic Studies*, 89(5): 2329–2380. [59]

Angelucci, Manuela, Dean Karlan, and Jonathan Zinman (2015). "Microcredit impacts: Evidence from a randomized microcredit program placement experiment by compartamos banco," *American Economic Journal: Applied Economics*, 7: 151–182. [4]

Angrist, Noam and Rachael Meager (2023). "Implementation matters: Generalising treatment effects in education," Working Paper BSG-WP-2023/054, Blavatnik School of Government, Working Paper Series. [4]

Attanasio, Orazio, Britta Augsburg, and Ralph De Haas (2018). "Microcredit Contracts, Risk Diversification and Loan Take-Up," *Journal of the European Economic Association*, 17(6): 1797–1842. [4]

Attanasio, Orazio, Britta Augsburg, Ralph De Hass, Emla Fitzsimons, and Heike Harmgart (2015). "The impacts of microfinance: Evidence from joint-liability lending in mongolia," *American Economic Journal: Applied Economics*, 7: 90–122. [4]

Attanasio, Orazio P. and Nicola Pavoni (2011). "Risk sharing in private information models with asset accumulation: Explaining the excess smoothness of consumption," *Econometrica*, 79(4): 1027–1068. [8]

Attanasio, Orazio P. and Miguel Székely (2004). "Wage shocks and consumption variability in Mexico during the 1990s," *Journal of Development Economics*, 73(1): 1–25. [2, 8]

Attanasio, Orazio P. and Guglielmo Weber (1989). "Intertemporal substitution, risk aversion and the euler equation for consumption," *The Economic Journal*, 99(395): 59–73. [59]

Attanasio, Orazio P. and Guglielmo Weber (1993). "Consumption growth, the interest rate and aggregation," *The Review of Economic Studies*, 60(3): 631–649. [59]

Augsburg, Britta, Ralph De Hass, Heike Harmgart, and Costas Meghir (2015). "The impacts of microcredit: Evidence from bosnia and herzegovina," *American Economic Journal: Applied Economics*, 7: 183–203. [4]

Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukerji, Marc Shotland, and Michael Walton (2017a). "From proof of concept to scalable policies; challenges and solutions, with an application," *Journal of Economic Perspectives*, 31(4): 73–102. [3, 36]

Banerjee, Abhijit, Emily Breza, Esther Duflo, and Cynthia Kinnan (2017b). "Do credit constraints limit entrepreneurship? heterogeneity in the returns to microfinance," Working Paper 17-104, Global Poverty Research Lab Working Paper Series. [4]

Banerjee, Abhijit, Esther Duflo, Rachel Glennerster, and Cynthia Kinnan (2015a). "The miracle of microfinance? evidence from a randomized evaluation," *American Economic Journal: Applied Economics*, 7: 22–53. [4]

Banerjee, Abhijit, Dean Karlan, and Jonathan Zinman (2015b). "Six randomized evaluations of microcredit: Introduction and further steps," *American Economic Journal: Applied Economics*, 7(1): 1–21. [1, 4]

Government of Bihar, The World Bank, the Global Facility for Disaster Reduction, and Recovery (2010). "Bihar kosi flood (2008) needs assessment report," Technical report. [12]

Bold, Tessa (2009). "Implications of endogenous group formation for efficient risk-sharing," *The Economic Journal*, 119: 562–591. [5]

Bold, Tessa, Mwangi Kimenyi, Germano Mwabu, Alice Ng'ang'a, and Justin Sandefur (2018).

"Experimental evidence on scaling up education reforms in kenya," *Journal of Public Economics*, 168: 1–20. [4]

Breza, Emily and Cynthia Kinnan (2021). "Measuring the Equilibrium Impacts of Credit: Evidence from the Indian Microfinance Crisis," *The Quarterly Journal of Economics*, 136(3): 1447–1497. [4]

Crépon, Bruno, Florencia Devoto, Esther Duflo, and WIlliam Parienté (2015). "Estimating the impact of microcredit on those who take it up: Evidence from a randomized experiment in morocco," *American Economic Journal: Applied Economics*, 7: 123–150. [4]

Datta, Upamanyu (2015). "Socio-economic effects of Jeevika: a large-scale self-help group project in Bihar, India," *World Development*, 68. [16]

Davis, Jonathan M.V., Jonathan Guryan, Kelly Hallberg, and Jens Ludwig (2017). "The economics of scale-up," Working Paper 23925, NBER Working Paper Series. [4]

Deaton, Angus and Christina Paxson (1994). "Intertemporal Choice and Inequality," *Journal of Political Economy*, 102(3): 437. [2, 8]

Demont, Timothée (2022). "Coping with shocks: How Self-Help Groups impact food security and seasonal migration," *World Development*, 105892. [4]

Dercon, Stefan (2004). "Growth and Shocks: Evidence from Rural Ethiopia," *Journal of Development Economics*, 74(2): 309–329. [59]

Dubois, Pierre, Bruno Jullien, and Thierry Magnac (2008). "Formal and informal risk sharing in ldcs: Theory and empirical evidence," *Econometrica*, 76(4): 679–725. [5]

Farhi, Emmanuel and Ivan Werning (2016). "Insurance and taxation over the life cycle," *Review of Economic Studies*, 83(2): 596–635. [59]

Feigenberg, Benjamin, Erica Field, and Rohini Pande (2013). "The economic returns to social interaction: Experimental evidence from microfinance," *The Review of Economic Studies*, 80(4 (285)): 1459–1483. [5]

Fischer, Greg (2013). "Contract structure, risk-sharing, and investment choice," *Econometrica*, 81(3): 883–939. [4]

Fisher, R.A. (1935). *The Design of Experiments*, Hafner Press. [16]

Genicot, Garance and Debraj Ray (2003). "Group formation in risk-sharing arrangements," *The Review of Economic Studies*, 70: 87–113. [5]

Government of Bihar (2011). "Annual report 2010-11," Technical report, Jeevika: Bihar Rural Livelihoods Promotion Society. [12]

Government of Bihar (2015). "Economic Survey 2014-15," Technical report, Finance Department. [10]

Government of Bihar (2016). "Economic survey 2015-16," Technical report, Finance Department. [12]

Government of India (2013). "Press Note on Poverty Estimates, 2011-12," Technical report, Planning Commission. [10]

Heß, Simon (2017). "Randomization inference with stata: a guide and software," *Stata Journal*, 17(3): 630–651. [16]

Hoffmann, Vivian, Vijayendra Rao, Vaishnavi Surendra, and Upamanyu Datta (2021). "Relief from usury: Impact of a self-help group lending program in rural India," *Journal of Development Economics*, 148(102567). [1, 4, 11, 13, 15, 16, 29, 55, 56, 57, 58, 62, 65, 67]

Holm, Sture (1979). "A simple sequentially rejective multiple test procedure," *Scandinavian Journal of Statistics*, 6(2): 65–70. [32]

Huggett, Mark (1993). "The risk-free rate in heterogeneous-agent, incomplete-insurance economies," *Journal of Economic Dynamics and Control*, 17(5–6): 953–969. [59]

Huggett, Mark, Gustavo Ventura, and Amir Yaron (2011). "Wealth distribution and risk sharing in a heterogeneous-agent model," *Journal of Economic Theory*, 146(2): 297–324. [59]

Imbens, Guido W. and Donald B. Rubin (2015). *Causal Inference in Statistics and the Social Sciences*, Cambridge and New York: Cambridge University Press. [16]

Imbens, Guido W. and Jeffrey M. Wooldridge (2009). "Recent developments in the econometrics of program evaluation," *Journal of Economic Literature*, 47(1): 5–86. [16]

Kaboski, Joseph P. and Robert M. Townsend (2012). "The impact of credit on village economies," *American Economic Journal: Applied Economics*, 4(2): 98–133. [4]

Kochar, Anjini, Bidisha Barooah, Chandan Jain, Geeta Singh, Nagabhushana Closepet, Raghunathan Narayanan, Ritwik Sarkar, and Rohan Shah (2020). "Impact evaluation of the national rural livelihoods project," Technical report, International Initiative for Impact Evaluation. [13]

Kochar, Anjini, Closepet Nagabhushana, Ritwik Sarkar, Rohan Shah, and Geeta Singh (2022). "Financial access and women's role in household decisions: Empirical evidence from india's national rural livelihoods project," *Journal of Development Economics*, 155(102821). [11]

Ligon, Ethan, Jonathan P. Thomas, and Tim Worrall (2002). "Informal Insurance Arrangements with Limited Commitment: Theory and Evidence from Village Economies," *The Review of Economic Studies*, 69(1): 209–244. [3, 5, 9, 10]

List, John (2024). "Field experiments: Here today gone tomorrow?" *The American Economist*, 69: 214–234. [3]

Meager, Rachael (2019). "Understanding the average impact of microcredit expansions: A bayesian hierarchical analysis of seven randomized experiments," *American Economic Journal: Applied Economics*, 11(1): 57–91. [1, 4]

Meager, Rachael (2022). "Aggregating distributional treatment effects: A bayesian hierarchical analysis of the microcredit literature," *American Economic Review*, 112: 1818–1847. [4]

Muralidharan, Karthik and Paul Niehaus (2017). "Experimentation at scale," *Journal of Economic Perspectives*, 31(4): 103–124. [3, 36]

Muralidharan, Karthik, Paul Niehaus, and Sandip Sukhtankar (2023). "General equilibrium effects of (improving) public employment programs: Experimental evidence from india," *Econometrica*, 91: 1261–1295. [3]

Olea, Jose Luis Montiel and Carolin Pflueger (2013). "A robust test for weak instruments," *Journal of Business and Economic Statistics*, 31(3): 358–369. [34, 63, 64]

Ravallion, Martin (1988). "Rural poverty and agriculture in côte d'ivoire," *Journal of Development Economics*, 28(2): 289–311. [59]

Rosenbaum, P (2002). *Observational Studies*, Springer. [16]

Tarozzi, Alessandro, Jaikishan Desai, and Kristin Johnson (2015). "The impacts of microcredit: Evidence from ethiopia," *American Economic Journal: Applied Economics*, 7: 54–89. [4]

Townsend, Robert M (1994). "Risk and Insurance in Village India," *Econometrica*, 62(3): 539–591. [2, 4, 6, 7, 8, 59]

Udry, Christopher (1994). "Risk and insurance in a rural credit market: An empirical investigation in northern nigeria," *The Review of Economic Studies*, 61: 495–526. [4]

Vivalt, Eva (2020). "How much can we generalize from impact evaluations?" *Journal of the European Economic Association*, 18: 3045–3089. [3, 4]

World Bank Group (2017). "A decadeof rural transformation: Lessons learnt from the bihar rural livelihoods project—jeevika," Technical report, World Bank. [12]

A Figures

Figure A1: *Jeevika's* phased roll-out and the experimental sample

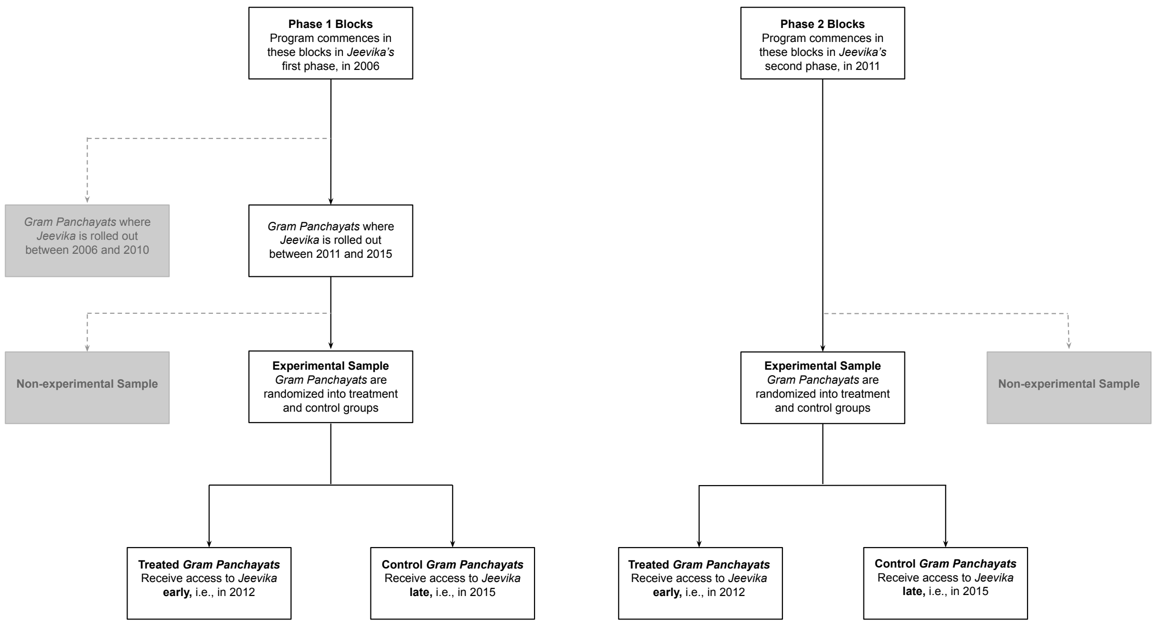
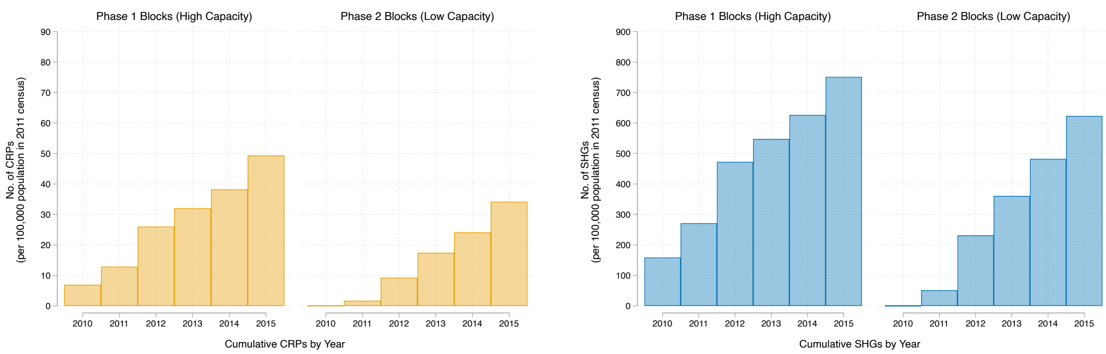
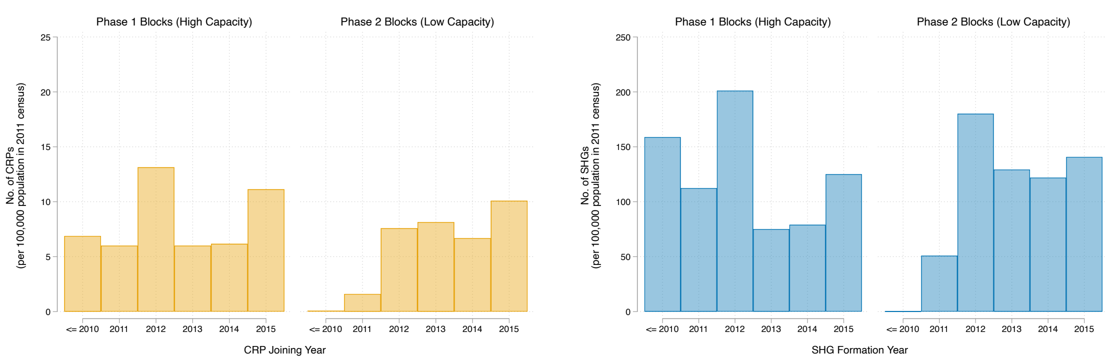


Figure A2: Cumulative program scale



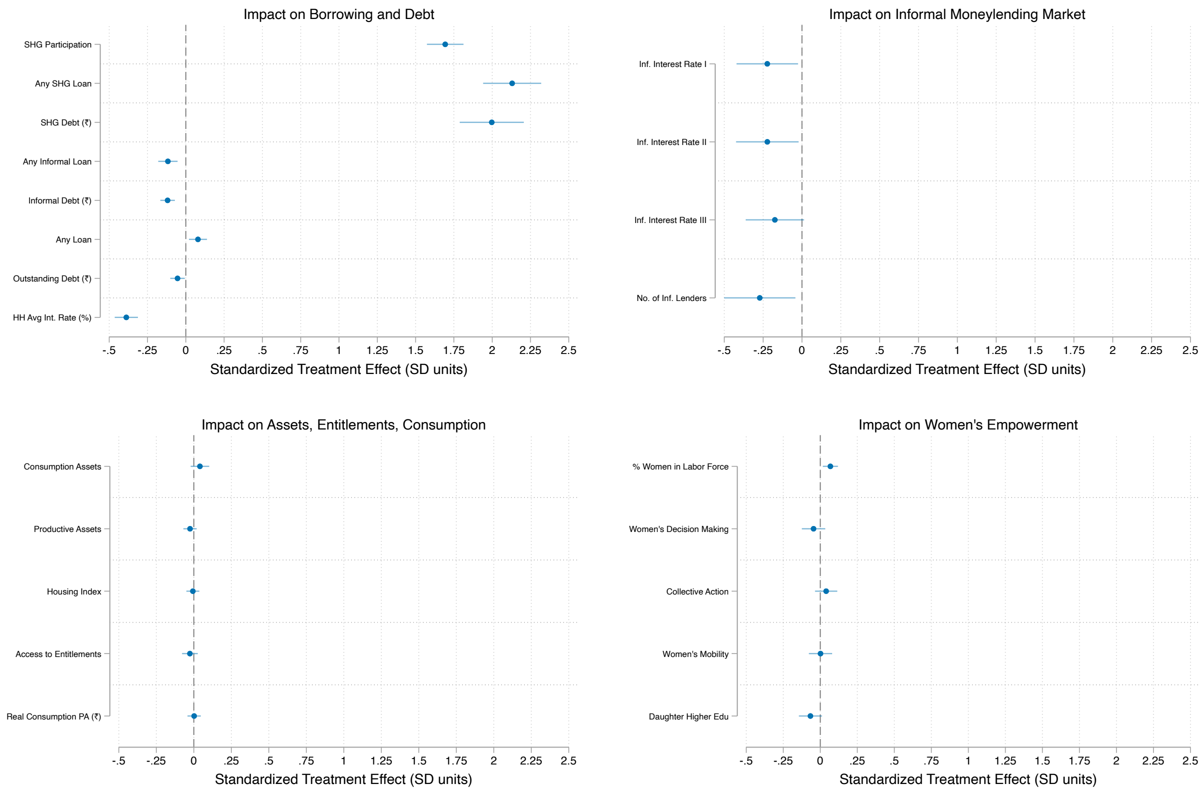
Source: *Jeevika* MIS data

Figure A3: Program scale added in each year



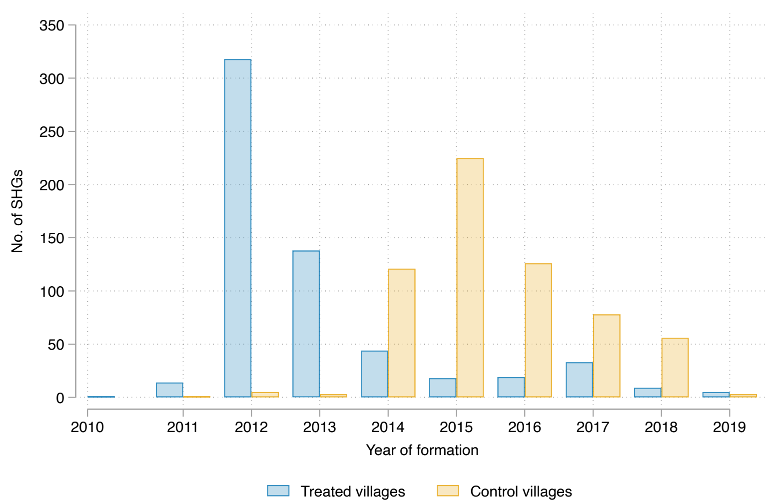
Source: *Jeevika* MIS data

Figure A4: *Jeevika's* impacts in control group standard deviation units



Notes: Outcomes are in control group standard deviation units.

Figure A5: SHGs formed in each year in 2019 SHG sample



Source: 2019 SHG survey data

B Additional Tables

Table B1: Baseline village characteristics across treatment and control villages

	Means			Normalized Differences [RI p-value]
	Obs	Control	Treatment	
	(1)	(2)	(3)	(4)
Village characteristics in 2001				
No. of households	333	612.805	632.720	0.041
				[0.684]
Population	333	3529.763	3670.780	0.051
				[0.608]
Dalit/Adivasi share	333	0.196	0.193	-0.020
				[0.780]
Female literacy rate	333	0.195	0.199	0.044
				[0.333]
Ratio of males to females in village	333	1.079	1.078	-0.013
				[0.572]
Infrastructure index	333	0.806	0.703	-0.064
				[0.630]
Village characteristics in 2011				
No. of households	333	897.254	940.262	0.059

				[0.616]
Population	333	4557.420	4811.683	0.070
				[0.537]
Dalit/Adivasi share	333	0.203	0.198	-0.042
				[0.610]
Female literacy rate	333	0.358	0.365	0.072
				[0.052]
Ratio of males to females in village	333	1.086	1.090	0.094
				[0.357]
Infrastructure index	333	0.677	0.762	0.053
				[0.489]
Village characteristics in baseline survey, 2011				
Variance of MPCE (All HHs, weighted)	333	0.120	0.114	-0.097
				[0.250]
Variance of MPCE (Dalit/Adivasi)	325	0.124	0.120	-0.064
				[0.755]
Variance of MPCE (non-Dalit/Adivasi)	294	0.120	0.119	-0.008
				[0.513]

Table B2: Baseline household characteristics across treatment and control villages

	Means			Normalized Differences [RI p-value]
	Obs	Control	Treatment	
	(1)	(2)	(3)	(4)
Household Characteristics				
Dalit/Adivasi?	8988	0.323	0.324	0.001
				[0.877]
Land-owner	8988	0.456	0.443	-0.027
				[0.658]
Household size	8988	5.463	5.476	0.005
				[0.885]
Female head?	8988	0.139	0.126	-0.038
				[0.320]
Any male migrants?	8988	0.437	0.404	-0.066
				[0.099]
Any SHG members?	8988	0.044	0.064	0.088
				[0.041]
Savings and Debt				
Any savings?	8988	0.383	0.422	0.079
				[0.156]
Outstanding debt (high cost)	8988	8.787	8.237	-0.037
				[0.509]

Outstanding debt (all)	8988	13.390	12.165	-0.064
				[0.195]
Outstanding debt (informal)	8988	11.085	10.773	-0.019
				[0.912]
Avg Int Rate (last year)	6460	5.088	5.167	0.055
				[0.277]
Avg Int Rate (informal, last year)	6389	5.100	5.178	0.056
				[0.265]
Consumption, Entitlements, Assets				
MPCE (all)	8970	764.183	751.619	-0.032
				[0.500]
Entitlements	8988	0.523	0.514	-0.017
				[0.654]
Consumer durables	8988	0.000	0.020	0.020
				[0.547]
Productive assets	8988	0.000	-0.050	-0.051
				[0.189]
Housing	8988	-0.000	-0.022	-0.022
				[0.943]
Women's Empowerment				
Women work proportion	8985	0.697	0.703	0.016
				[0.850]
Women decision making	8988	0.835	0.867	0.091

				[0.053]
Collective action	8988	0.813	0.810	-0.008
				[0.846]
Women's mobility	8960	0.326	0.318	-0.023
				[0.696]
Daughter higher education	5235	0.377	0.366	-0.023
				[0.675]
Attrition				
Attrition	8988	0.028	0.029	0.008
				[0.749]

Table B3: Village characteristics across program phases

	(1) Obs	(2) Phase 1 Mean (SD)	(3) Phase 2 Mean (SD)	(4) Difference in Means
Village characteristics in 2001				
All villages in study blocks				
No. of households	1419	573.096	327.931	-239.694**
		(558.484)	(420.831)	(90.132)
Population	1419	3282.147	1941.201	-1309.394**
		(3200.922)	(2390.628)	(488.926)
Dalit/Adivasi share	1419	0.183	0.251	0.067
		(0.132)	(0.211)	(0.052)
Female literacy rate	1419	0.157	0.249	0.093***
		(0.078)	(0.123)	(0.026)
Ratio of males to females in village	1419	1.092	1.059	-0.033**
		(0.069)	(0.119)	(0.014)
Infrastructure index	1419	0.446	0.393	-0.048
		(1.571)	(1.724)	(0.478)
Village characteristics in 2011				
All villages in study blocks				
No. of households	1419	863.418	460.188	-395.157***
		(848.477)	(581.185)	(129.024)
Population	1419	4344.902	2447.376	-1856.510***
		(4203.645)	(2991.539)	(614.669)
Dalit/Adivasi share	1419	0.193	0.259	0.065
		(0.144)	(0.221)	(0.055)
Female literacy rate	1419	0.333	0.423	0.091***
		(0.086)	(0.119)	(0.026)
Ratio of males to females in village	1419	1.098	1.086	-0.013
		(0.134)	(0.129)	(0.014)
Infrastructure index	1419	0.198	0.585	0.399
		(1.559)	(1.728)	(0.605)
Village characteristics in baseline survey, 2011				
Study villages				
Village-level variance of MPCE (Weighted)	333	0.124	0.104	-0.020**
		(0.064)	(0.059)	(0.009)
Village-level variance of MPCE (only Dalit/Adivasi HHs)	325	0.126	0.113	-0.013
		(0.060)	(0.060)	(0.011)
Village-level variance of MPCE (only non-Dalit/Adivasi HHs)	296	0.126	0.105	-0.021*
		(0.082)	(0.078)	(0.010)
No. of villages	333	109	224	
Treated villages	164	51	113	
Control villages	169	58	111	

Notes: Infrastructure index is the first principal component from a principal components analysis of the following infrastructure measures: whether the there is a bank in the village, whether there is a credit society in the village, whether the village has pucca roads, whether the village has domestic electricity connections, whether the village has agricultural electricity connections, whether the village has rail connectivity, whether the village has bus connectivity, whether the village has access to tap water, whether the village has any wells, whether the village has any tubewells, whether the village has any educational facilities from primary school onwards, whether the village has any medical facilities.
Robust standard errors clustered at the block level in parentheses in column 4.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B4: Household characteristics across program phases in 2011 baseline survey

		Means		Difference
	Obs	Phase I Blocks	Phase II Blocks	in Means
Household characteristics				
Dalit/Adivasi?	8988	0.282	0.282	-0.000
		(0.010)	(0.007)	(0.041)
Land-owner	8988	0.450	0.473	0.023
		(0.014)	(0.010)	(0.029)
HH Size	8988	5.621	5.432	-0.189*
		(0.074)	(0.047)	(0.111)
Female Head?	8988	0.142	0.126	-0.016
		(0.009)	(0.007)	(0.019)
Any Male Migrants?	8988	0.373	0.441	0.069**
		(0.014)	(0.010)	(0.032)
Any SHG Members?	8988	0.098	0.031	-0.067***
		(0.009)	(0.003)	(0.020)
Savings and debt				
Any Savings?	8988	0.467	0.372	-0.095**
		(0.014)	(0.010)	(0.045)
Outstanding Debt (High Cost)	8988	8.707	8.500	-0.207
		(0.524)	(0.304)	(0.902)
Outstanding Debt (All)	8988	13.439	12.849	-0.590

		(0.595)	(0.413)	(1.070)
Outstanding Debt (Informal)	8988	12.010	10.698	-1.311
		(0.570)	(0.323)	(0.902)
Avg Int Rate (last year)	6460	4.884	5.198	0.314***
		(0.037)	(0.033)	(0.105)
Avg Int Rate (informal, last year)	6389	4.912	5.201	0.289***
		(0.036)	(0.033)	(0.105)
Consumption, assets, entitlements				
Monthly per capita Cons Exp (MPCE, ₹)	8970	674.084	801.851	127.767***
		(9.960)	(8.684)	(21.793)
Access to entitlements?	8988	0.461	0.526	0.064
		(0.014)	(0.010)	(0.039)
Consumer Durables Index	8988	0.176	-0.036	-0.212***
		(0.031)	(0.020)	(0.069)
Productive Assets Index	8988	-0.191	0.087	0.278***
		(0.021)	(0.024)	(0.062)
Housing Quality Index	8988	0.203	-0.100	-0.303***
		(0.033)	(0.021)	(0.089)
Women's roles and capabilities				
Women Work Proportion	8985	0.653	0.711	0.059*
		(0.012)	(0.009)	(0.031)
Women Decision Making	8988	0.942	0.808	-0.133***

		(0.007)	(0.008)	(0.028)
Collective Action	8988	0.846	0.789	-0.057**
		(0.010)	(0.008)	(0.024)
Women's Mobility	8960	0.378	0.291	-0.087***
		(0.011)	(0.007)	(0.021)
Daughter Higher Education	5235	0.437	0.357	-0.080**
		(0.019)	(0.013)	(0.036)
Attrition				
Attrition	8988	0.032	0.028	-0.004
		(0.005)	(0.003)	(0.005)
No. of HHs	8988	2,920	6,068	
Treated HHs	4,472	1,413	3,059	
Control HHs	4,516	1,507	3,009	

Robust standard errors clustered at the cluster (CLF) level in parentheses in column 4.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B5: Is the variance of consumption growth zero (in control group villages)?

Variable	Phases	Mean	t-stat	p-value	N
Dalit/Adivasi Households	1 & 2	0.254	28.981	0.000	164
	1	0.250	18.440	0.000	55
	2	0.255	22.630	0.000	109
Non-*Dalit/Adivasi* Households	1 & 2	0.255	20.106	0.000	153
	1	0.228	11.820	0.000	53
	2	0.269	16.431	0.000	100

T-tests of variance of consumption growth means against zero in control group villages.

Table B6: Impact of *Jeevika*'s institutional capacity on village risk sharing

	Variance of change in log consumption			
	(1)	(2)	(3)	(4)
Dalit/Adivasi Households				
Treatment	0.004	0.003	0.010	0.007
	(0.017)	(0.016)	(0.018)	(0.017)
Initial program scale	0.013***	0.012**	0.018*	0.012
	(0.004)	(0.005)	(0.011)	(0.013)
Treatment × Initial program scale	-0.019***	-0.020***	-0.046***	-0.040***
	(0.007)	(0.006)	(0.015)	(0.015)
Initial program scale squared			-0.001	-0.000
			(0.001)	(0.002)
Treatment × Initial program scale squared			0.005**	0.004**
			(0.002)	(0.002)
Controls	No	Yes	No	Yes
Obs	325	325	325	325
Clusters	50	50	50	50
Mean (Control Group)	0.254	0.254	0.254	0.254
Non-*Dalit/Adivasi* Households				
Treatment	-0.008	-0.005	-0.001	0.002
	(0.022)	(0.021)	(0.023)	(0.024)
Initial program scale	-0.010**	-0.010	0.012	0.021
	(0.005)	(0.007)	(0.013)	(0.021)
Treatment × Initial program scale	-0.011	-0.017**	-0.043**	-0.049**
	(0.009)	(0.008)	(0.018)	(0.021)
Initial program scale squared			-0.004**	-0.004*
			(0.001)	(0.002)
Treatment × Initial program scale squared			0.006**	0.006*
			(0.002)	(0.003)
Controls	No	Yes	No	Yes
Obs	294	294	294	294
Clusters	49	49	49	49
Mean (Control Group)	0.255	0.255	0.255	0.255

Notes: The outcome is variance of change in log consumption. Initial program scale is defined in Section 4.2. The variance measure is computed relying on sampling weights to reconstitute caste composition of the village. Regressions in even columns include controls (from the 2001 census) for the village Dalit and Adivasi population share, village female literacy rate, village sex ratio, an index of village infrastructure. Robust standard errors clustered at the program cluster level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B7: Impact of *Jeevika*'s institutional capacity on village risk sharing
(Alternate measures of initial program scale)

	Variance of change in log consumption									
	Initial scale is CRPs per 1,000 hectares to be mobilized		Initial scale is SHGs per 100 target at scale		Initial scale is CRPs per 1,000 target at scale		Initial scale is SHGs per 1,000 target, 2011 census		Initial scale is CRPs per 10,000 target, 2011 census	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment	-0.004	0.013	-0.005	-0.002	-0.006	0.010	-0.006	-0.002	-0.006	0.012
	(0.014)	(0.017)	(0.015)	(0.015)	(0.015)	(0.017)	(0.015)	(0.016)	(0.015)	(0.017)
Initial program scale	0.002	0.035**	-0.006	0.025	0.002	0.074*	-0.015	0.061	0.011	0.196*
	(0.007)	(0.016)	(0.007)	(0.036)	(0.011)	(0.037)	(0.018)	(0.101)	(0.034)	(0.106)
Treatment × Initial program scale	-0.017**	-0.064***	-0.035**	-0.098*	-0.101**	-0.320**	-0.101**	-0.320**	-0.086	-0.372***
	(0.007)	(0.019)	(0.014)	(0.052)	(0.017)	(0.028)	(0.043)	(0.129)	(0.052)	(0.097)
Initial program scale squared		-0.005***		-0.009		-0.019**		-0.070		-0.162**
		(0.002)		(0.009)		(0.008)		(0.074)		(0.073)
Treatment × Initial program scale squared		0.009***		0.032		0.036***		0.319*		0.354***
		(0.003)		(0.025)		(0.009)		(0.169)		(0.100)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs	333	333	333	333	333	333	333	333	333	333
Clusters	50	50	50	50	50	50	50	50	50	50
Mean (Control group)	0.243	0.243	0.243	0.243	0.243	0.243	0.243	0.243	0.243	0.243

Notes:The outcome is the variance of change in log consumption, and is computed relying on sampling weights to reconstitute caste composition of the village. All odd columns present results from regressions of the outcome on an indicator for treatment status interacted with initial scale, while all even columns present results from regressions of the outcome on an indicator for treatment status interacted with initial scale and the indicator for treatment status interacted with initial scale squared. All regressions include controls (from the 2001 census) for the village Dalit and Adivasi population share, village female literacy rate, village sex ratio, an index of village infrastructure.
Initial scale definitions: (i) no. of CRPs in 2011 (prior to experiment) per 1000 hectares over which mobilization is yet to occur in the cluster (excluding control group), used in columns (1) and (2); (ii) no. of SHGs in 2010 (prior to experiment) per 100 women to be mobilized in the cluster to reach reach target at scale (excluding control group), used in columns (3) and (4); (iii) no. of CRPs in 2011 (prior to experiment) per 1,000 women to be mobilized in the cluster to reach reach target at scale (excluding control group), used in columns (5) and (6); (iv) no. of SHGs in 2010 (prior to experiment) per 100 women to be mobilized in the cluster (excluding control group) based on 2011 census, used in columns (7) and (8); (v) no. of CRPs in 2011 (prior to experiment) per 1,000 women to be mobilized in the cluster (excluding control group) based on 2011 census, used in columns (9) and (10).
Robust standard errors clustered at the program cluster level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B8: Impact of *Jeevika*'s institutional capacity on household savings and consumption

	Any savings? (%)	Consumption Asset Index	MPCE	Cons Growth
	(1)	(2)	(3)	(4)
Treatment	27.017***	0.012	0.025	0.037
	(2.616)	(0.047)	(0.021)	(0.030)
Initial program scale	-0.382	-0.020	0.013**	0.017**
	(1.204)	(0.020)	(0.006)	(0.008)
Treatment × Initial program scale	-1.316	0.030	-0.007	-0.020
	(1.186)	(0.024)	(0.010)	(0.014)
Obs	8987	8987	8800	8800
Clusters	50	50	50	50
Mean (Control group)	46.698	0.000	6.796	0.486

Notes: Outcome in col (1) is in percentage points. Outcomes in cols (2)-(4) are standardized using the mean and standard deviation for households in the control group in clusters with zero initial program scale. Initial program scale is defined in Section 4.2. All regressions include strata fixed effects, household-level baseline controls from Hoffmann et al. (2021), and village-level controls from the 2001 census (Dalit/Adivasi population share, female literacy rate, sex ratio, infrastructure index). Robust standard errors clustered at the program cluster level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C Impact of Initial Program Scale on Outcomes of Interest in Hoffmann et al. (2021)

Table C1: Impact of *Jeevika*'s institutional capacity on program take-up, debt and borrowing

	Any SHG Member? (%)	Borrowed in the last year? (%)			Outstanding Debt (in '000 2011 ₹)			Interest Rate (HH avg)
		SHG	Informal	All	SHG	Informal	All	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	1.779***	2.124***	-0.102**	0.085*	1.955***	-0.116***	-0.046	-0.393***
	(0.083)	(0.163)	(0.049)	(0.044)	(0.164)	(0.032)	(0.030)	(0.055)
Initial program scale	0.037	0.012	0.051***	0.030***	-0.050	0.019**	0.027**	-0.023
	(0.036)	(0.043)	(0.014)	(0.010)	(0.062)	(0.008)	(0.013)	(0.028)
Treatment × Initial program scale	-0.086	0.020	0.001	0.006	0.058	-0.006	-0.011	0.013
	(0.053)	(0.081)	(0.022)	(0.018)	(0.083)	(0.014)	(0.016)	(0.028)
Obs	8851	8987	8987	8987	8987	8987	8987	6805
Clusters	50	50	50	50	50	50	50	50
Mean (Control group)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: This table presents heterogeneity by initial program scale results (direct effects of *Jeevika*, Table 2 in Hoffmann et al. (2021)) in control group standard deviation units. Initial program scale is defined in Section 4.2. All regressions include strata fixed effects, household-level baseline controls from Hoffmann et al. (2021) and village-level controls from the 2001 census (Dalit/Adivasi population share, female literacy rate, sex ratio, infrastructure index). Robust standard errors clustered at the program cluster level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C2: Impact of *Jeevika*'s institutional capacity on the informal credit market

	Household Survey Unweighted		Household Survey Weighted by Loan Size		Village Survey	
	Interest Rate	Interest Rate (Selection Corrected)	Interest Rate	Interest Rate (Selection Corrected)	Interest Rate	No. of Informal Lenders
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.041	-0.117	-0.036	-0.114	-0.140	-0.330**
	(0.047)	(0.089)	(0.047)	(0.090)	(0.125)	(0.147)
Initial program scale	-0.043*	-0.030	-0.037	-0.024	-0.038	0.039
	(0.023)	(0.102)	(0.023)	(0.112)	(0.076)	(0.059)
Treatment × Initial program scale	0.034	0.016	0.030	0.011	-0.035	0.068
	(0.025)	(0.047)	(0.024)	(0.047)	(0.071)	(0.071)
Obs	6211	6211	6211	6211	327	333
Clusters	50	50	50	50	50	50
Mean (Control group)	0.000	0.000	0.000	0.000	0.000	0.000

Notes: This table presents heterogeneity by initial program scale results (effects of *Jeevika* on the informal credit market, Table 3 in Hoffmann et al. (2021)) in control group standard deviation units. Initial program scale is defined in Section 4.2. All regressions include strata fixed effects, household-level baseline controls from Hoffmann et al. (2021) and village-level controsl from the 2001 census (Dalit/Adivasi population share, female literacy rate, sex ratio, infrastructure index). Robust standard errors clustered at the program cluster level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C3: Impact of *Jeevika*'s institutional capacity on household asset position, entitlements, and welfare

	Consumption Asset Index	Production Asset Index	Housing Asset Index	Access to Entitlements?	Monthly Consumption PA
	(1)	(2)	(3)	(4)	(5)
Treatment	0.021	-0.008	-0.027	-0.042	-0.004
	(0.048)	(0.036)	(0.027)	(0.032)	(0.027)
Initial program scale	-0.019	0.003	-0.048**	0.020**	0.010
	(0.023)	(0.008)	(0.019)	(0.009)	(0.010)
Treatment × Initial program scale	0.016	-0.019	0.007	0.026*	0.004
	(0.026)	(0.014)	(0.015)	(0.013)	(0.011)
Obs	8987	8987	8987	8987	8987
Clusters	50	50	50	50	50
Mean (Control group)	0.000	0.000	0.000	0.000	0.000

Notes: This table presents heterogeneity by initial program scale results (effects of *Jeevika* on household asset position, entitlements, and welfare, Table 4 in Hoffmann et al. (2021)) in control group standard deviation units. Initial program scale is defined in Section 4.2. All regressions include strata fixed effects, household-level baseline controls from Hoffmann et al. (2021) and village-level controsl from the 2001 census (Dalit/Adivasi population share, female literacy rate, sex ratio, infrastructure index). Robust standard errors clustered at the program cluster level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C4: Impact of *Jeevika*'s institutional capacity on women's economic roles, empowerment, and aspirations

	Proportion HH Women Working	Women Decision Making	Collective Action	Women's Mobility	Daughter Education
	(1)	(2)	(3)	(4)	(5)
Treatment	0.070*	-0.041	0.020	0.007	-0.043
	(0.037)	(0.065)	(0.055)	(0.051)	(0.048)
Initial program scale	0.016	-0.000	-0.019	0.049***	0.027
	(0.014)	(0.015)	(0.026)	(0.016)	(0.031)
Treatment × Initial program scale	0.013	0.013	0.023	-0.003	-0.032
	(0.021)	(0.025)	(0.030)	(0.026)	(0.035)
Obs	8830	8841	8841	8813	4090
Clusters	50	50	50	50	50
Mean (Control group)	0.000	0.000	0.000	0.000	0.000

Notes: This table presents heterogeneity by initial program scale results (effects of *Jeevika* on women's economic roles, empowerment, and aspirations, Table 5 in Hoffmann et al. (2021)) in control group standard deviation units. Initial program scale is defined in Section 4.2. All regressions include strata fixed effects, household-level baseline controls from Hoffmann et al. (2021) and village-level controsl from the 2001 census (Dalit/Adivasi population share, female literacy rate, sex ratio, infrastructure index). Robust standard errors clustered at the program cluster level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D Functional form assumptions

In this Appendix, we discuss the assumptions we make to define deviations from risk sharing and its relationship with SHG quality.

D.1 CRRA and alternative functional form assumptions.

Standard test in the perfect risk sharing literature follow the Townsend (1994), which involves the dynamics of the marginal utility of consumption. The assumption of power utility, with Constant Relative Risk Aversion (CRRA), which we have used, is central to deriving these equations that are linear in logs. However, some alternatives have been used in the literature. A widely used utility function has been the Constant Absolute Risk Aversion (CARA) specification, which implies log utility linear in the level of consumption. While CRRA and CARA preferences are analytically simple, several authors have examined risk sharing under utility specifications that depart from these assumptions. An interesting direction in the present context is one that incorporates subsistence or threshold consumption effects, of the Stone Geary type. Examples include Ravallion (1988); Dercon (2004). In these settings, a consumption floor, k, alters insurance incentives and typically predicts higher volatility of consumption among poorer households—a feature often observed in village-level data.[12]

D.2 Does SHG quality improve risk sharing?

Section 6.2 established that higher initial program scale leads to better-functioning SHGs, with larger savings pools and stronger governance. We now test whether this improved SHG quality is the channel through which program scale affects risk sharing.

[12]The assumptions of power utility and expected utility maximization can be relaxed in other directions. One interesting example are models with recursive preferences that separate risk aversion from intertemporal substitution motives, such as Attanasio and Weber (1989, 1993); Farhi and Werning (2016); Huggett et al. (2011). In addition to changes to the hypotheses, one can consider different market and risk share settings. Huggett (1993) and Ambrus et al. (2022) consider risk sharing in incomplete markets and networks.

The theoretical framework in Section 2 implies that risk sharing depends on the value of continued group membership relative to the outside option. Households honor informal agreements when the expected future benefits of membership, access to current and future group resources, exceed the short-term gain from deviation. Put differently, households insure other members against unexpected income shortfalls only when they expect reciprocity in the future, and this expectation depends on the group continuing to function well. Higher-quality SHGs raise the value of membership by accumulating larger pools of loanable funds and maintaining governance practices that signal continued access to resources; these same features also reflect members' willingness to honor risk-sharing arrangements.

As shown in Section 6.2, SHGs formed under higher initial program scale accumulated larger pools of loanable funds per member, and exhibited stronger organizational performance. If program scale improves risk sharing by producing higher-quality SHGs, then measures of SHG quality should mediate the relationship between scale and risk sharing. However, SHG quality is endogenous to household choices and may be correlated with unobserved determinants of risk sharing. So, we test this by instrumenting SHG quality with the interaction between treatment status and initial program scale, the same source of variation that generated the reduced-form results in Section 5.3.

Measuring SHG quality. The analysis in Section 6.2 used SHG-level outcomes measured in 2019. Here, we require village-level measures of SHG quality at the end of 2014, the end of the period over which we measure consumption growth. We construct three proxies for the value of SHG membership. The first is accumulated loanable funds per household in the village, constructed using monthly loanable funds per member (from the 2019 SHG survey, averaged 2016–2018) multiplied by membership and months of operation for each SHG formed by end-2014 (from MIS data). We aggregate to the village level and normalize by 2011 census household counts. The second is the village-level share of SHGs with high *panchsutra* adherence, assuming that the 2019 village average from the SHG survey reflects the 2014 average. The third is a standardized index averaging the first two measures.

Each measure relies on assumptions: that the 2019 loanable funds accumulation rates are representative of 2014, and that *panchsutra* scores persist over time. We interpret all three as noisy proxies for the same latent variable, the value of SHG membership, and consistent results across measures would suggest we are capturing the underlying construct rather than idiosyncratic measurement error.

Empirical strategy. We estimate a two-stage least squares specification relating village-level risk sharing, measured by the village-level variance of consumption growth between 2011 and 2014 to average SHG quality in a village or the value of SHG membership, $\mathbf{V}_{vgc}$. SHG quality is the endogenous regressor and is instrumented by the interaction between treatment status and initial program scale, $\mathbf{T}_{gc} \times \mathbf{Scale}_c$. The first stage projects SHG quality onto the instrument and controls; the second stage relates risk sharing to predicted SHG quality.

The first-stage equation is:

$$\mathbf{V}_{vgc} = \delta_1 \mathbf{T}_{gc} \times \mathbf{Scale}_c + \delta_2 \mathbf{T}_{gc} + \delta_3 \mathbf{Scale}_c + \mathbf{X}'_{vgc} \delta_4 + \mathbf{S} + u_c \tag{11}$$

where $\mathbf{V}_{vgc}$ is the relevant SHG quality outcome for a village v, in GP g, and cluster c; $\mathbf{T}_{gc}$ is the GP's treatment assignment; $\mathbf{Scale}_c$ is initial program scale at the program-cluster level; $\mathbf{X}_{vgc}$ is a vector of pre-treatment village-level measures from the 2001 census used by the government to assign regions to program phases (share of *Dalit*/*Adivasi*) households, sex ratios, female literacy rate, infrastructure index); $\mathbf{S}$ are randomization strata fixed effects. Robust standard errors are clustered at the program-cluster level.

The second-stage equation is:

$$Var_{vgc}(\triangle_{2011}^{2014} \log c^{ivgc}) = \gamma_1 \widehat{\mathbf{V}}_{vgc} + \gamma_2 \mathbf{T}_{gc} + \gamma_3 \mathbf{Scale}_c + \mathbf{X}'_{vgc} \gamma_4 + \mathbf{S} + \epsilon_c \tag{12}$$

where, $Var_{vgc}(\triangle_{2011}^{2014} \log c^{ivgc})$ is the variance of consumption growth in the village between

2011 and 2014, for a village, v, in gram panchayat, g, in cluster, c; $\widehat{\mathbf{V}}_{vgc}$ is the predicted SHG quality outcome from the first-stage regression. Again, robust standard errors are clustered at the program-cluster level.

Exclusion restriction and interpretation. The IV specification uses the same source of variation as the reduced-form analysis, the interaction between treatment and initial program scale, now as an instrument for SHG quality. The exogeneity of this instrument was established in Section 4.2.

We now discuss potential violations of the exclusion restriction. First, if initial program scale reduced informal interest rates, this could directly impact household borrowing as self-insurance. However, Table C2 shows that the interaction between treatment and initial program scale does not significantly affect interest rates, suggesting this channel is not operative. Second, the instrument might affect risk sharing through income or wealth effects. Table B8 shows no significant effect of initial program scale on mean consumption or consumption growth in treated areas, and Table C1, Table C3, and Table C4 show no effects on other household outcomes documented in Hoffmann et al. (2021). Third, SHGs provide members with access to savings, credit, and stronger social networks. As noted in Section 5.3, these are not threats to identification, but rather the mechanisms through which SHGs can affect risk sharing. Consistent with this interpretation, Table 6 shows that higher initial program scale reduces borrowing in response to shocks, which we interpret as reflecting improved informal insurance.

We also interpret our three measures of SHG quality (loanable funds, panchsutra adherence, and the composite SHG quality index) as noisy proxies for the same underlying latent variable, i.e., the value of group membership. The exclusion restriction therefore requires only that the instrument affects risk sharing through this single channel, even though we measure it imperfectly. Consistent results across measures suggest that we are capturing the underlying construct rather than idiosyncratic measurement error.

Finally, many household-level impacts representing alternative channels may themselves be downstream of, or closely related to, SHG quality, making them difficult to separate empirically. We therefore view the IV results as suggestive evidence of the proposed mechanism. The reduced-form result, that higher initial program scale improves risk sharing among treated villages, remains identified and policy-relevant regardless of the precise channel.

Table C5: Impact of village SHG quality on village risk-sharing (IV)

	Variance of change in log consumption					
	Endogenous regressor is *Loanable Funds*		Endogenous regressor is *High Panchsutra Score*		Endogenous regressor is *SHG Quality Index*	
	(1)	(2)	(3)	(4)	(5)	(6)
	Second-stage					
Loanable Funds	-0.096*	-0.113**				
	(0.050)	(0.053)				
Panchsutra Score			-0.206*	-0.251*		
			(0.120)	(0.129)		
SHG Quality Index					-0.053*	-0.068*
					(0.032)	(0.035)
	First-stage co-efficients					
Initial program scale × Treatment	0.146***	0.145***	0.068**	0.066**	0.263***	0.241***
	(0.052)	(0.054)	(0.027)	(0.027)	(0.100)	(0.092)
Controls	No	Yes	No	Yes	No	Yes
Obs	333	333	333	333	333	333
Clusters	50	50	50	50	50	50
Olea and Pflueger (2013) F-statistic	7.869	7.108	6.395	6.096	6.855	6.887
Critical Values (% worst case bias)						
τ=5%	37.418	37.418	37.418	37.418	37.418	37.418
τ=10%	23.109	23.109	23.109	23.109	23.109	23.109
τ=20%	15.062	15.062	15.062	15.062	15.062	15.062
τ=30%	12.039	12.039	12.039	12.039	12.039	12.039
Anderson-Rubin F-stat	5.218	7.825	5.218	7.825	5.218	7.825
Anderson-Rubin p-value	0.027	0.007	0.027	0.007	0.027	0.007

Notes: The **excluded instrument** is Treatment × Initial program scale. Initial program scale is defined in Section 4.2. The **outcome** is the variance of change in log consumption, and is computed relying on sampling weights to reconstitute caste composition of the village. In columns (1) and (2), the **endogenous regressor** is the *village-level accumulated per-household loanable funds with all SHGs ('00,000) at the end of 2014*. In each village, the total per-household loanable funds at the end of 2014 is constructed using the mean monthly per-member loanable funds (from member savings and grants) across all SHGs in a village from the 2019 SHG survey, the number of members in each SHG, the number of SHGs and the age of SHGs at the end of 2014 from the MIS data. In columns (3) and (4), the **endogenous regressor** is the *share of SHGs in the village with a high panchsutra score (at least 4 out of 5)*. This assumes that the panchsutra scores at the end of 2014 is the same as in 2019. In columns (5) and (6), the **endogenous regressor** is the *mean SHG quality in a village*. SHG quality is a standardized index constructed as the mean of the standardized quality variables in the 2019 SHG survey: (i) accumulated village loanable funds per household; (ii) share of SHGs with high panchsutra scores in the village. The first-stage coefficient rows report the coefficients on the excluded instrument in the first-stage regression (Treatment × Scale), where the endogenous regressor is regressed on the instrument and controls. **Controls or included instruments** are: treatment, program scale, and in even numbered columns, village-level controls from the 2001 census for Dalit and Adivasi population shares, female literacy rate, sex ratio, and an infrastructure index. Weak identification is assessed using the Olea-Flueger F-statistic, which accounts for heteroskedastic errors. (Olea and Pflueger, 2013). Robust standard errors clustered at the program cluster level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Results. Table C5 presents the IV estimates. The second-stage coefficients on SHG quality are negative across all proxies for quality and all specifications, indicating that higher SHG quality improves risk sharing (or reduces the variance of consumption growth). The effect is statistically significant at the 5% level in column (2) and at the 10% level in the remaining columns. The first-stage results confirm that the interaction between treatment and initial program scale significantly predicts all three measures of SHG quality. This supports the hypothesis that program scale operates through SHG quality, i.e., the value of SHG membership.

The Olea-Pflueger effective F-statistics (ranging from 6.1 to 7.9) fall below even the most lenient critical values reported in Table C5, so we cannot rule out substantial worst-case bias in the two-stage least squares point estimates (Olea and Pflueger, 2013). For this reason, we base inference on the Anderson-Rubin test, whose size is correct regardless of instrument strength. The Anderson-Rubin test rejects the null that the instrument has no effect on risk sharing ($p = 0.007$ with controls); the sign and significance of the effect are therefore robust to the weak first stage, even though the magnitudes are not reliably estimated. This inference mirrors the reduced-form results and does not by itself isolate the mechanism. We therefore interpret these IV results as suggestive evidence that SHG quality is the channel linking program scale and risk sharing, consistent with the participation-constraint mechanism, while acknowledging that the weak first stage limits the precision of our magnitude estimates.

When the value of continued SHG membership, i.e., the resources a group commands and the expectation that it will keep functioning, tightens participation constraints, an intervention that raises the quality of such groups can improve risk sharing. Our findings provide suggestive evidence of exactly this margin, which, to our knowledge, has received limited direct empirical attention. SHGs can thus play an important role in insuring households against idiosyncratic shocks, but only when they function well, moving villages closer to full risk sharing.

E Data Appendix

This section describes the construction of the main variables used in the paper. Data comes from (1) household and village surveys conducted in 2011 and 2014 by Hoffmann et al. (2021); (2) SHG surveys conducted in 2019; (3) India's 2001 census and 2011 census; (4) *Jeevika's* Management Information System (MIS).

1. **Village risk sharing**: This variable is constructed using the household consumption module from the Hoffmann et al. (2021) study. The primary outcome is the village-level variance of the change in log real monthly per capita consumption expenditure (MPCE, in 2011 rupees) between the 2011 baseline and 2014 endline.

   Both survey waves utilized an identical consumption module covering food and household goods (30-day recall) as well as education, medical, and clothing expenses (1-year recall). Of the 73 total items surveyed, we exclude 8 items due to missing or unrealistic recorded values (4 at baseline and 4 at endline), leaving 65 items for the final measure. To minimize the influence of outliers, MPCE is winsorized at the 1st and 99th percentiles, calculated separately for treatment and control groups at endline, prior to log-transformation.

   We calculate the village-level variance for the full sample (with sampling weights) and for two distinct sub-samples: (1) Dalit/Adivasi households and (2) non-Dalit/Adivasi households. Sub-sample variances are computed without sampling weights. Following the Hoffmann et al. (2021) study's stratified design (70% Dalit/Adivasi; 30% non-Dalit/Adivasi), we employ inverse probability weights to reconstitute the actual caste composition of each village. These weights are normalized to sum to one at the village level, ensuring each village carries equal weight in the final analysis.

2. **Loanable funds per member (SHG-level)**: Total savings accumulated by members over the life of the SHG (mean monthly savings rate averaged over April 1, 2016 to

December 31, 2018, multiplied by months in existence) plus all grants received through December 31, 2018, divided by the number of members.

3. **Loanable funds per household (village-level)**: The IV specifications in the appendix require a village-level measure aligned with the 2014 risk-sharing data. For each SHG formed by end-2014 (identified from MIS data), we multiply monthly loanable funds per member (from the 2019 SHG survey, averaged 2016–2018) by membership and months of operation. We aggregate across SHGs to the village level and normalize by household counts from the 2011 Census.

4. **High panchsutra (SHG-level)**: The panchsutra is scored on five criteria: (1) regular meetings, (2) regular saving, (3) regular lending, (4) regular repayment, and (5) up-to-date record-keeping. We construct an indicator for each criterion as follows:

   - Regular meetings: SHG met at least 4 times per month in the last year (consistent with prescribed weekly meetings)
   - Regular saving: Members saved at least ₹480 per month in the last year
   - Regular lending: At least one loan taken by members in the last year
   - Regular repayment: SHG imposes penalties for default (no direct measure of repayment schedules is available)
   - Record-keeping: All of cash book, general ledger, minutes book, loan ledger, and savings ledger are maintained

   High panchsutra equals 1 if at least 4 of these 5 criteria are met.

5. **High panchsutra share (village-level)**: The IV specifications in the appendix require a village-level measure aligned with the 2014 risk-sharing data. We calculate the share of SHGs in each village with high panchsutra adherence, assuming the 2019 village average from the SHG survey reflects the 2014 average.

6. **SHG quality index (shg-level)**: A standardized index averaging loanable funds per member and the high panchsutra indicator.

7. **SHG quality index (village-level)**: A standardized index averaging loanable funds per household and the high panchsutra share.

8. **Baseline controls from Hoffmann et al. (2021)**: Any household participation in SHGs; any household savings; real outstanding high-cost debt (in 2011 ₹); average interest rate on loans in the preceding year; productive asset index; consumption asset index; housing index; real monthly consumption per adult equivalent; any access to entitlements; proportion of women working in the household; women's decision-making index; any collective action undertaken; preference for daughter's higher education; women's mobility index; Dalit/Adivasi indicator; landless indicator (all measured at baseline).