NBER WORKING PAPER SERIES

DOING BETTER AT DOING GOOD: LESSONS FROM A BAYESIAN META-ANALYSIS AFTER 20 YEARS OF POVERTY GRADUATION PROGRAMS

Samantha Carter
Edward Jee
Dean Karlan
Patrick Premand

Working Paper 35991
http://www.nber.org/papers/w35991

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2026

We thank Makayla Barker, Riza Halili, Carlie Lamke, and Giuliana Triberti for excellent research assistance; Julius Rüschenpöhler for his partnership in creating DEvidence and the data encoding, processing, quality checks, and organization needed to assemble the data; the Partnership for Economic Inclusion for the data underlying the evidence review for the 2021 State of Economic Inclusion Report; participants in the PEI Impact Collaborative Technical workshop and Technical Members' meeting for comments; and, the countless researchers, implementers, and donors for the studies and knowledge that have made this work possible. We are grateful for comments on an earlier draft from Diva Dhar, John Loeser, David McKenzie, Adam Salisbury, Victoria Strokova and a World Bank DIME reviewer. We used AI tools for some assistance with coding and the creation of some figures and tables, as well as light copy-editing on parts of the manuscript. We reviewed all content and are responsible for any errors. All opinions are ours and do not necessarily represent the views of the World Bank, its affiliated organizations, its Executive Directors, the governments it represents, or the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2026 by Samantha Carter, Edward Jee, Dean Karlan, and Patrick Premand. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Doing Better at Doing Good: Lessons from a Bayesian Meta-Analysis After 20 Years of Poverty Graduation Programs

Samantha Carter, Edward Jee, Dean Karlan, and Patrick Premand

NBER Working Paper No. 35991
November 2026
JEL No. C11, D61, H53, I32, I38, O12, O13

ABSTRACT

Multifaceted poverty Graduation programs that aim to increase income generating activities for low-income households are being rapidly adopted and scaled by nonprofit organizations and governments. This paper conducts a Bayesian hierarchical meta-analysis of 49 studies (encompassing 81 treatment variations) on poverty Graduation programs, defined as containing at least three of five components: consumption support, savings facilitation, skills training, coaching, and a lumpsum transfer. The analysis finds large gains across economic and psychosocial outcomes: monthly consumption rises by 38 USD PPP (11 percent), income by 28 USD PPP (15 percent), alongside improvements in savings, assets, food security, mental health, and women's empowerment. Beyond these average treatment effects, three results stand out. First, programs with large transfers (at least 400 USD PPP) have larger treatment effects but are not more cost-effective. Second, more expensive programs generate larger treatment effects, but the improvements are not commensurate with the cost increases; thus, lower cost programs tend to be more cost-effective. Third, programs implemented from 2016 onward perform considerably better than those implemented prior to 2016: benefit-cost ratios from 2016 onward are 5.39, compared to 2.89 through 2015. Neither changes in cost nor in high-level design (components) explain this upward trend. Considering other heavily studied interventions, the results show a similar upward trend for treatment effects on behavior change from financial education but not for unconditional cash transfers. The paper hypothesizes a mechanism of continuous sector-community learning and improved implementation quality.

Samantha Carter
Livelihood Impact Fund
samanthaw.carter@gmail.com

Edward Jee
The University of Chicago
edjee@uchicago.edu

Dean Karlan
Northwestern University
and NBER
dean.karlan@gmail.com

Patrick Premand
The World Bank
ppremand@worldbank.org

1 Introduction

Graduation programs aim to generate long-run socioeconomic impacts by helping low-income households engage in new or expanded income generating activities. Under the premise that market frictions are not single-faceted, the programs deliver a package of components, specifically some combination of consumption support, savings facilitation, skills training, coaching and an asset transfer or a lump-sum cash grant.

Tests of an initial design by BRAC in Bangladesh (Bandiera et al., 2017) and seven coordinated adaptations in Ethiopia, Ghana, Honduras, India, Pakistan, Peru, and the Republic of Yemen (Banerjee et al., 2015a; Brune et al., 2022) led to an initial "proof of concept" of the Graduation model. Economic inclusion programs, including Graduation approaches, have been further adopted, adapted and scaled by nonprofit organizations and governments in 88 countries, reaching over 70 million individuals (Arévalo-Sánchez et al., 2024). In the process, implementers have experimented with both simpler and more complicated versions, while also pushing to find ways to deliver programs at a lower unit cost without sacrificing treatment effects. The rapid spread of poverty Graduation programs, often also considered within a broader set of "economic inclusion" programs that may not include all core components, has been accompanied by a robust research agenda, testing everything from alternative bundles of components to tailoring for particular populations to varying scale and delivery approaches. With 41 journal articles, reports, and working papers covering randomized evaluations of the Graduation model, we can now begin to quantitatively aggregate results and examine questions about optimal design and trends in the practice and impact of these programs.

We conduct a meta-analysis of multi-faceted programs that include at least three of the five core Graduation program components: (i) consumption support, (ii) training, (iii) coaching, (iv) savings facilitation or encouragement of savings, and (v) an asset or lump sum cash grants. We therefore cover both "core" Graduation programs, as initially implemented by NGOs such as BRAC, as well as a broader set of "economic inclusion" programs which are inspired by the Graduation model and keep its multifaceted logic, but are, in some cases, lighter in scope. This inclusion criterion reflects the heterogeneity of programs now operating around the world (Arévalo-Sánchez et al., 2024), while maintaining sufficient conceptual coherence to permit meaningful pooling. In total, the meta-analysis draws on 49 study records from 41 publications, reports, and working papers, covering 40 unique randomizations. These contain 81 reported treatment variations and 123 treatment-variation-wave observations; consolidating repeated reports of the same intervention lineage yields 68 unique treatment arms and 109 canonical treatment-arm-wave observations.

We build on two recently conducted meta-analyses that encompass Graduation programs. Kondylis and Loeser (2021) examine the persistence of consumption impacts for 17 RCTs, comparing Graduation programs to unconditional cash transfers. Results support the idea that Graduation programs address constraints beyond mere liquidity: Graduation programs generate more persistent impacts, but are less cost-effective in the short run and show more heterogeneous impacts – which plausibly reflects their

increased complexity. Leight et al. (2024), casting a wider net by including non-randomized studies as well as "cash-plus" interventions (which include Graduation studies but also studies with only one added component beyond cash), examines 104 studies and impacts focused on consumption, income and labor supply. They find larger effects for "cash-plus" interventions than mere cash, as well as higher cost-effectiveness under some assumptions.[1]

We differentiate from these meta-analyses on four dimensions. First, we focus on Graduation programs, including neither unconditional cash transfers nor "cash-plus" interventions that have only one additional component beyond cash.[2] Second, beyond documenting gains on monthly consumption (by 38 USD PPP, or 11%) and income (by 28 USD PPP, or 15%), we also find broader improvements in savings, assets, employment, self-employment, food security, mental health, life satisfaction, aspirations, women's agency, and empowerment. Third, we attempt to draw inferences about the importance of specific components. Programs with large transfers (at least 400 USD PPP) have larger treatment effects. However, an important nuance is that they do not have larger effect per unit of cost. More broadly, more expensive programs generate larger treatment effects, but lower cost programs tend to be more cost-effective. Aside from larger transfers, the data do not tell a consistent pattern whereby any other component stands out as either necessary or irrelevant.

Finally, we examine impacts for programs implemented more recently: average treatment effects, cost-effectiveness, and benefit-cost ratios are all higher for programs in the past ten years, relative to the first ten years when Graduation programs were implemented. Specifically, for example, using the treatment effect persistence observed from our cross-study analysis, the benefit-cost ratio increases from 2.89 for programs implemented through 2015 to 5.39 for programs implemented since 2016. This contrasts with a common expectation in policy evaluation that effects attenuate as programs move from initial pilots to wider implementation (Allcott, 2015; Al-Ubaydli et al., 2017; Banerjee et al., 2017; List, 2022). We compare these patterns to two other types of interventions to assess hypotheses for this trend: unconditional cash transfers (which acts as a placebo, on the premise that cash is less sensitive to implementation fidelity but similarly prone to various manifestations of publication bias) and financial education (which acts as a robustness test, for the opposite premise in terms of potential sensitivity to implementation quality). We find a flat trend for unconditional cash transfers, but a similarly upward trend for financial education impact on behavior change. We discuss possible mechanisms but see this primarily as evidence for improved productivity via operational efficiency (although recognize that this comparison to other sectors is not definitive). More broadly, this pattern also builds on the productivity growth and knowledge spillovers literature, i.e., how diffusion of knowledge and practices across producers can drive productivity gains (Bloom et al., 2013; Crespi et al., 2008).

[1] More recently, Baseler et al. (2026) focuses on economic and psychological spillover effects, including an examination of important methodological questions regarding the method and public or private nature of the randomization that determined program participation.

[2] The exception is for the benefit-cost analysis, when we compare our results to a re-analysis of 114 unconditional cash transfer studies in Crosta et al. (2026).

2 Study Sample and Data

2.1 Study Search and Inclusion Criteria

We built the study sample through a staged screening process, summarized in Figure A.1. The initial search used the terms graduation, livelihoods and multifaceted, together with randomized, evaluation, poverty or low-income in Google Scholar, EconLit, and the AEA RCT Registry.[3] The search yielded 3,391 database records, which we cross-checked for exclusion errors against 115 studies from the PEI State of Economic Inclusion Report evidence database (Andrews et al., 2021). After removing duplications, the search produced 2,604 unique candidate records. We then applied an initial relevance screen, retaining 306 semi-finalist studies for detailed eligibility review. The large difference between these two steps is due to the search terms (especially "graduation") picking up many studies on education that are not relevant to poverty graduation programs. We then assessed whether each intervention satisfied our graduation program definition of including at least three of five core components: a lump-sum transfer, a stream of smaller transfers for consumption support, savings support, curriculum-based training, and coaching or follow-up support.[4]

Specifically, each study we include meets the following criteria:

1. The study is a randomized controlled trial.
2. The comparison arm received no intervention, a placebo, or business-as-usual services (such as small regular transfers for consumption support).
3. The intervention was administered in a low- or middle-income country (as defined by World Bank classification in August 2025).

At the level of the treatment arm, we use the following inclusion criteria:

1. The treatment arm administered at least three of the following five eligible treatment components:
   (a) Consumption support
   (b) Savings support (access or encouragement)
   (c) Asset or lump sum cash transfer
   (d) Skills training
   (e) Coaching
2. The study reports results on any of the following outcomes: consumption, income, assets, employment, savings, food security, mental health, life satisfaction, aspirations, or female empowerment.

[3] Specifically, a June 2024 EconLit search used `("graduation" OR "livelihood" OR "multifacet*") AND ("randomized" AND "evaluation") AND ("poverty" OR "low income") NOT "teacher"`, with related terms and all years enabled.

[4] In September 2026, we added 8 studies that had been issued since our initial search and met the search criteria, with the objective of ensuring that the results remained up to date.

Our inclusion criteria imply that we focus on multifaceted graduation and economic inclusion interventions. We do not include simple "cash-plus" programs that administer a cash transfer and only one additional treatment component (such as training). We note that some recent government-led programs layer multifaceted interventions on top of public works programs. In this case, workfare replaces regular cash transfers for consumption support. We do not include these programs unless they independently meet the three-component criterion. This is because the number of studies is limited at the time of writing, and public works imply different types of labor shifts and opportunity costs than regular consumption support.

2.2 Sample Description

The final analysis sample draws on 49 study records from 41 publications, reports, and working papers, covering 40 unique randomizations. Across these records, we observe 81 reported treatment variations and 123 treatment-variation-wave observations; consolidating repeated reports of the same intervention lineage yields 68 unique treatment arms and 109 canonical treatment-arm-wave observations. Figure 1 summarizes the final analysis sample.[5] Our sample includes programs evaluated in 26 countries. Coverage is concentrated in Sub-Saharan Africa, but it spans South Asia, Southeast Asia, Latin America, and the Middle East. About a quarter of the studies occurred in fragile or post-conflict settings (specifically: Afghanistan, Burkina Faso, Chad, the Democratic Republic of Congo, Niger, South Sudan, Uganda (with refugees from the DRC), and the Republic of Yemen).

Appendix Table A.1 reports component support and outcome sample sizes. In terms of components, all treatment arms include training, and a majority includes an asset or lump-sum transfer, coaching, and/or savings support. Despite these high-level similarities, the component details vary across treatment arms. The value of lump-sum transfers varies sharply across treatment arms: some variants omit a productive transfer entirely, while others provide transfers worth more than USD 1,000 in 2020 PPP terms (Data Appendix Table 2). Program cost and duration also vary meaningfully throughout the sample. Income, consumption, and assets are all reported for a majority of treatment arms, with coverage falling sharply for other outcomes, as we discuss next.

2.3 Data Extraction

2.3.1 DEVIDENCE Ecosystem for Data Extraction and Quality Control

We extracted treatment effects from published papers and appendices using the DEVIDENCE extraction workflow from the Global Poverty Research lab at Northwestern University. For each study, the extraction instrument recorded study metadata, publication and appendix links, treatment-control comparisons, survey waves, units of randomization and analysis, treatment and control sample sizes, treatment components, outcome labels, units, transformations, baseline values, endline values, and

[5] Data Appendix Table 1 lists all the included studies and references. Data Appendix Table 2 provides details on the program components for each unique treatment arm.

reported treatment-effect estimates. We use reported estimates rather than re-estimating effects from microdata.

When multiple estimates were available, we prioritized estimates that were disaggregated by treatment arm and survey round. Within each arm-wave cell, we selected the measure with the highest control group mean when there were multiple measures of monetary and non-index outcomes; for indexed or standardized outcomes, we used the broadest measure possible. For example, the highest-control-mean selection criterion led us to prioritize total household consumption over food consumption alone and total household income over income from a single activity.

2.3.2 Key Outcome Variables

We report 12 outcomes grouped into consumption, income, food security, savings, assets, employment, psychological well-being, and female empowerment.

Consumption

Monthly household consumption captures household-level consumption or total expenditure. If not available at the household level, we scale per-capita consumption linearly by household size. If total consumption is not available, we use the consumption or expenditure category with the next-largest control group mean.

Income

Monthly household income captures household-level income or earnings, and total revenue for businesses. Similarly to consumption, if household-level measures are not available, we proxy by per-capita income adjusted for household size. If no aggregate measure of total income is available, we use the income, earnings, or category with the next-largest control group mean.

Food Security

Food security captures study-reported food security scores, indices, and related indicators, coded so that higher values indicate greater food security. The included studies use both established instruments—including the Food Consumption Score (FCS), Food Insecurity Experience Scale (FIES), Household Food Insecurity Access Scale (HFIAS), and Household Hunger Scale (HHS)—and author-designed indices based on experiences such as insufficient food, meal skipping, reduced portions, or going an entire day without eating. We prioritize author-reported standardized indices where available and harmonize the direction of all included measures.

Savings

Savings stock prioritizes savings balances or stocks where available. When no stock measure is reported, we use savings flows, such as deposits over the month preceding follow-up.

Non-Financial Assets

Total assets picks up measures of the total stock of household assets. If not available, we use the asset category with the next-largest control group mean.

Employment

Any work reported captures whether households report any form of work. We use this as the reported employment outcome because activity type and work-intensity measures are less comparable across studies. We note that studies do not consistently report engagement across all forms of work (in either wage or self-employment, on or off-farm). *Self-employment/Business operation* captures whether households report self-employment in off-farm income-generating activities or business operation.

Psychological Well-Being

Mental health captures mental health and psychological well-being indices, such as the Kessler mental health score. If no standard index is reported, we use author-designed scores. In the absence of any general index, we prioritize depression scores and indices. If none of the above are available, we use single mental health symptoms or perceived stress scores. In 6 of the 18 studies contributing to the mental health outcome, the selected measure is a narrower symptom, stress, or psychological-distress scale rather than a broad mental health index. *Satisfaction and happiness* comprises simple, one-item life satisfaction and perceived happiness measures. *Aspirations and expectations* is comprised mostly of measures of general, educational, or child-related aspirations or general measures of future expectations. If neither is available, we proxy with measures of optimism and whether respondents expect life to improve. In three studies contributing to aspirations and expectations, the selected measure is a proxy based on optimism, hopelessness, or whether respondents expect life to improve.

Empowerment

To capture notions of women's empowerment, we use agency and general empowerment indices. *Agency* is a measure of household-level decision-making power and mostly picks up on indices about women's say in resource allocation. If no index is reported, we proxy by simple one-item measures of power in household resource allocation and decision-making. *General empowerment* is strictly a measure of general women's or partner's empowerment or disempowerment indices.

2.4 Costs Data

Of the 68 unique treatment arms in our sample, 48 provide intervention cost estimates, which for descriptive reporting we express in constant 2020 market-exchange-rate USD per household, using reported nominal costs where available and PPP-backconverted nominal equivalents otherwise.[6] Across these 48 arms in Data Appendix Table 3, estimated costs range from 102 USD to 3,090 USD per

[6]The quality of cost data varies widely, with some papers mentioning a unit cost without providing detailed information and others providing detailed cost breakdowns.

participating household. The median cost is 445 USD, and the mean is 819 USD. One notable pattern is that several of the higher-cost programs and pilots were designed before 2016.[7] Government adoption of graduation programs increased over time and accelerated after 2016, and programs with government involvement tend to have lower costs when they leverage national social protection systems.

Partly due to the varying cost reporting standards, we estimate parallel cost models with the same observable program characteristics: a nominal model for funder-facing cost reporting and a PPP model for analyses that compare costs directly with PPP treatment effects. Among the 38 canonical treatment-arm observations used to estimate the nominal unit-cost model, the covariates explain substantial within-sample variation in costs (adjusted $R^2 = 0.41$, see Appendix Table A.2). Given the selected sample of programs reporting costs, we use the models for prediction rather than interpreting individual coefficients causally. The cost-effectiveness analyses below use predictions from the PPP model so that costs and treatment effects share the same basis; reported PPP costs remain the denominator in the benefit-cost analysis.

3 Methodology

3.1 Hierarchical Models and Program Components

Synthesizing evidence on graduation programs presents three challenges: separating true contextual heterogeneity from sampling variation, describing how impacts vary with program composition, and deriving persistence and cost-benefit estimates from uncertain impacts. These challenges are particularly important when investigating graduation programs that differ across settings, bundle multiple costly components, and are designed to generate effects that may persist beyond the observed follow-up period.

Recent syntheses of cash and cash-plus interventions similarly pool experimental estimates to examine how intervention size, program scope, and follow-up timing shape livelihood impacts and their persistence (Kondylis and Loeser, 2021; Leight et al., 2024). Our framework shares this emphasis on heterogeneity and persistence, while focusing specifically on graduation programs and propagating uncertainty through Bayesian models of program components, persistence, and cost-effectiveness.

We study these questions within a Bayesian hierarchical framework. Following the standard random-effects approach to evidence synthesis, we treat reported estimates as noisy measurements of heterogeneous underlying effects (Rubin, 1981). Bayesian hierarchical versions of this approach have increasingly been used to synthesize randomized evaluations in development economics (Meager, 2019; Vivalt, 2020; Crosta et al., 2026). We first use the model to separate cross-study heterogeneity from sampling noise, and then extend it to examine variation by program components and follow-up

[7] The original pilots in Ethiopia, Ghana, Pakistan, and Peru (Banerjee et al., 2015a) all began implementation in or before 2011; Bedoya et al. (2019) study a component of a larger World Bank program that began in 2015; and the WINGS program studied by Blattman et al. (2016) was implemented beginning in 2009. The Graduating to Resilience program studied by Beam et al. (2026) was designed with and supported by USAID, with higher costs relative to programs funded through private philanthropy.

timing and to estimate persistence and cost-effectiveness.

Let $\hat{\tau}_i$ denote the reported treatment-effect estimate for study-arm-wave observation i, and let s_i denote its reported standard error. We treat $\hat{\tau}_i$ as a noisy estimate of a latent treatment effect τ_i and model the latent effects as draws from a common distribution:

$$\hat{\tau}_i \mid \tau_i, s_i \sim \mathcal{N}(\tau_i, s_i^2) \tag{1}$$

$$\tau_i \mid \mu_\tau, \sigma_\tau \sim \mathcal{N}(\mu_\tau, \sigma_\tau^2) \tag{2}$$

The parameter μ_τ is the pooled average treatment effect for a given outcome, while σ_τ^2 measures true variation in effects across study arms and follow-up waves after accounting for reported sampling error.

The Gaussian structure is well suited to aggregating randomized evaluations reported in the economics literature. The source studies typically report regression-based point estimates and standard errors that are justified by large-sample normal approximations. The model in Equation (2) uses that same approximation at the first stage, while adding a second stage that allows true effects to differ across observations. Classical fixed-effect meta-analysis is nested as the limiting case where $\sigma_\tau^2 = 0$, but the hierarchical model allows the data to indicate when that restriction is inappropriate.

Because graduation interventions differ substantially in their design, we extend the baseline model to ask whether program components explain part of the observed heterogeneity. A graduation program is not a single homogeneous treatment: programs vary in transfer size, intensity, and the inclusion of components such as consumption support, savings facilitation, training, and coaching. Follow-up timing also varies. We therefore model latent treatment effects as a function of study-arm-level covariates X_i:

$$\tau_i \mid X_i, \beta, \sigma_u \sim \mathcal{N}(X_i\beta, \sigma_u^2) \tag{3}$$

The vector X_i contains program characteristics such as whether all five core components are present, the size of the transfer, whether the program includes savings support, and whether coaching and training are present separately or jointly. The residual variance σ_u^2 captures heterogeneity that remains after accounting for these observed design features. For the main heterogeneity comparison, we fit the no-covariate baseline and component specification to the same complete-case arm–follow-up observations for each outcome. All eligible follow-ups are retained, so the reported dispersion is arm-wave heterogeneity. We report residual and predicted-mean dispersion and the component model's within-model heterogeneity share; the unrestricted pooled-effect and cost-benefit analyses retain their own analysis samples.

The component models are best interpreted as descriptive evidence on program heterogeneity rather than as causal estimates of individual components. Most graduation evaluations do not independently randomize coaching, training, savings support, or transfer size within the same study. These components are bundled with each other and with contextual features such as delivery modalities, target

population, market access, and local implementation capacity. The component regressions therefore summarize how impacts covary with program design across the evidence base, while stopping short of identifying the marginal causal effect of adding any one component. Additional robustness checks are presented in appendix and compare the implementation-year, era, and persistence specifications with matched-sample models adding a shared randomization-level intercept.[8] These sensitivities permit shared latent effects within experiments while retaining conditionally independent sampling errors; but they do not supply the sampling-error covariance induced by shared controls or repeated respondents.

Estimating σ_τ^2 and σ_u^2 also gives the analysis a direct measure of external validity. If residual heterogeneity is small, estimates from one study can be pooled strongly with evidence from other studies. If residual heterogeneity is large, the model places wider bounds on what can be inferred for an unstudied context. This partial-pooling logic shrinks noisy isolated estimates toward the model-implied mean, but the degree of shrinkage is learned from the data rather than imposed mechanically.

3.2 Modeling Persistence of Impacts

Graduation programs are designed to alter households' long-run economic trajectories, so the timing of follow-up measurement is key. The theory of the graduation model implies that a large initial transfer, coupled with training and coaching, may generate returns that persist or compound after the program ends. At the same time, assets may depreciate, business opportunities may fade, and temporary consumption gains may decay. A single follow-up estimate cannot distinguish these possibilities.

We study persistence by extending the covariate model to include follow-up timing. For income and consumption, we annualize treatment effects and estimate models of the form

$$\tau_i \mid t_i, \beta_0, \beta_t, \sigma_u \sim \mathcal{N}(\beta_0 + \beta_t t_i, \sigma_u^2) \tag{4}$$

where t_i is years since the start of the program or baseline. The coefficient β_t summarizes the relationship between follow-up timing and treatment effects across the available evidence.

The persistence estimates should be interpreted as descriptive time-pattern evidence rather than as pure within-study decay parameters. Follow-up timing is not randomly assigned across studies, and long-run observations are available for only a subset of programs and contexts. We also estimated robustness specifications in restricted timing windows to test whether conclusions are driven by very short-run follow-ups, by a small number of long-run studies, or by a broader pattern across the evidence base. We summarize these models using posterior intervals for β_t and the posterior probability $\Pr(\beta_t > 0)$.

[8] Table A.13 contains the results.

3.3 Bayesian Cost-Benefit Analysis

Impact estimates alone are often insufficient for policy evaluation. Graduation programs typically combine large transfers with sustained staff engagement, training, coaching, and monitoring, all of which add to costs. As a result, evidence of positive effects on consumption or income is not by itself enough to establish policy viability. The estimated benefits must also be evaluated relative to program costs.

For each study arm with available cost data, we translate posterior impact draws into benefit-cost ratios using the observed sequence of follow-ups before projecting benefits beyond the final observed survey. Let C_a denote the per-household cost of program arm a, let T_a be the last observed follow-up year for that arm, and let $\tau^{(m)}_{a,w(t)}$ denote posterior draw m of the annualized monetized treatment effect from the observed wave mapped to year t. We discount at an annual rate of $r = 0.05$ for every program, both implementation eras, and the cash-transfer comparison. For the permanent-persistence and linear-decay benchmarks, the draw-level value of benefits at the last observed follow-up and the benefit-cost ratio are

$$B^{(m)}_a = \sum_{t=1}^{T_a} (1+r)^{T_a - t} \tau^{(m)}_{a,w(t)} + \tau^{(m)}_{a,T_a} \phi \tag{5}$$

$$\mathrm{BCR}^{(m)}_a = \frac{B^{(m)}_a}{C_a (1+r)^{T_a}} \tag{6}$$

Here $w(t)$ maps each year to the nearest observed follow-up wave, and ϕ is the terminal continuation factor applied to the last observed consumption effect. The summation uses all available follow-up evidence rather than discarding intermediate waves; in these two benchmarks, only benefits after the last observed follow-up are projected. The posterior distribution of $\mathrm{BCR}^{(m)}_a$ is the object we summarize in the cost-benefit analysis tables, including posterior means or medians, posterior intervals, and the probability that benefits exceed costs, $\Pr(\mathrm{BCR}_a > 1)$. Both $\tau^{(m)}_{a,w(t)}$ and C_a are expressed in constant USD with 2020 PPP, so benefits and costs enter the ratio with the same unit. When we report cost levels for a funder-facing interpretation, we instead display constant 2020 nominal USD, using reported nominal costs where available and PPP-backconverted nominal equivalents otherwise.

Our monetized benefit measure uses posterior estimates of consumption gains, since consumption is widely measured and directly interpretable as household economic welfare. We do not add asset or savings stocks to the benefit measure because doing so may double-count transferred assets or subsequent accumulation already reflected in consumption. Non-monetary outcomes, including empowerment, mental health, aspirations, and food security, are analyzed separately and are not monetized in the benefit-cost calculation.

The Bayesian framework carries uncertainty from the impact estimates into the final return calculation. Rather than computing a single deterministic ratio from point estimates, we calculate benefit-cost ratios using posterior draws of the relevant treatment effects. This produces a posterior distribution of returns for each program arm, allowing us to summarize not only the expected benefit-cost ratio

but also uncertainty about whether benefits exceed costs.

Persistence assumptions are central to the cost-benefit calculation because they determine the projected benefit path. We therefore report three alternative scenarios. The first, which we call the "linear-decay scenario," applies a path based on a long-run follow-up of a graduation program in Ethiopia (Barker et al., 2024), with benefits declining by 12.5% per year and reaching zero after eight years. The second, which we call the "estimated-persistence scenario," uses each arm's first eligible follow-up as its anchor and evaluates years 1–10 on a common horizon. It adds the pooled follow-up slope times the distance from the anchor year, back-casting before the anchor and forecasting after it, and floors every evaluated annual effect at zero before discounting at 5 percent, using aforecasting assumption based on the association estimated in Section 3.2.[9] The third scenario, "permanent persistence," follows the terminal-value logic of the cost-benefit analysis in Banerjee et al. (2015a): the final observed annual consumption effect persists forever, beginning one year after the last observed follow-up, so $\phi = \sum_{k=1}^{\infty}(1+r)^{-k} = 1/r$.

3.4 Bayesian Estimation and Inference

Bayesian estimation is useful in this setting because the number of graduation evaluations is limited relative to the heterogeneity we wish to model. Estimating cross-site variance and multiple component associations from a limited evidence base risks overfitting if the model is allowed to chase every observed difference in point estimates. We therefore use weakly regularizing priors, trading a small amount of prior structure for more stable estimates of average effects, heterogeneity, component associations, and persistence parameters.

We scale the priors to the empirical dispersion of treatment effects within each outcome. This makes the amount of regularization comparable across outcomes measured in different units, such as dollars, binary indicators, and standardized indices, without forcing effects onto a common substantive scale. For the program component models and persistence models, the coefficient priors are centered at zero, so the prior benchmark is no association between a component and impacts, and no systematic time trend.

We estimate the models using Hamiltonian Monte Carlo in Stan. Hierarchical models create posterior dependence between latent treatment effects, average effects, heterogeneity parameters, and covariate coefficients. Hamiltonian Monte Carlo simulates a particle moving through the posterior surface, using local slopes to make long, directed proposals that follow high-probability regions rather than taking short random-walk steps.

The resulting posterior distributions are summarized using posterior means or medians and 95% posterior intervals. Throughout the paper, these intervals should be read as Bayesian posterior intervals rather than frequentist confidence intervals. They represent uncertainty after combining the reported

[9]The Appendix provides additional robustness checks. For instance, Appendix Table A.14 holds the horizon and sample fixed while varying the slope and zero floor. Appendix Table A.15 separates within- and between-arm timing associations.

treatment-effect estimates, their standard errors, the hierarchical model, and the priors described above.[10]

4 Results

4.1 Average Treatment Effects

4.1.1 Economic Well-Being

Figure 2 presents forest plots displaying estimates for consumption and income for each treatment arm for which they are observed. Forest plots for all outcomes are reported together in Appendix Figure A.2, Panels A–L.[11] Figure 3 displays average treatment effects for all the outcomes for all programs, along with their 95% credible intervals and median control means.

Consumption is often considered the main economic well-being outcome for Graduation programs and is also one of the most often measured (52 treatment arms; 83 arm-waves). The posterior average treatment effect shows an increase in monthly household consumption of +38.1 USD (2020 PPP), or 11% relative to the median control mean of 361 USD (2020 PPP).[12] Figure 2 shows that there is meaningful heterogeneity across treatment arms: point estimates range from modestly negative to above 100 USD (2020 PPP) per month, although the negative estimates are imprecise.

The largest point estimates were generated by "full" Graduation interventions including all five components, e.g. World Vision's Graduation program in Somalia (Leight et al., 2025), the Graduating to Resilience Program in Uganda with either individual or group coaching (Beam et al., 2026), the Afghanistan Targeting the Ultra Poor program (Bedoya et al., 2019), as well as the WINGS program in Uganda that did not provide consumption support (Blattman et al., 2016). On the other hand, small and not statistically significant consumption gains were found for "streamlined" arms in Burkina Faso (Bossuroy et al., 2024), Kenya (Gobin et al., 2016) and India (Bauchet et al., 2015), as well as for "full" pilot interventions in Yemen (Brune et al., 2022), South Sudan (Chowdhury et al., 2016), and Honduras (Banerjee et al., 2015a). The Honduras pilot suffered from severe implementation issues, and intervention delivery in Burkina Faso, Yemen and South Sudan was affected by instability and other operational challenges (Bossuroy et al. (2024), Banerjee et al. (2015a)).[13] Note that the programs in Somalia (Leight et al., 2025) and Afghanistan (Bedoya et al., 2019) that had some of the largest treatment effects on consumption were also implemented in fragile contexts, but both were delivered

[10] Appendix Tables A.21 and A.20 report the model-specific priors, estimation-sample scale parameters, and sampling diagnostics.

[11] Where multiple follow-ups are available, separate wave-specific estimates and credible intervals are displayed within each treatment-arm row. We analyze the time dimension and persistence of impact in detail in Section 4.4.1 below.

[12] See Appendix Table A.3 for the estimates.

[13] In the Honduras pilot, 83 percent of participants selected chickens as their productive asset. A large fraction of the chickens died due to illness by the time of the second endline survey. In Burkina Faso, the cash grants that served as asset transfers were delayed during COVID-19, and field teams' contracts ended around the time the grants were finally delivered, so that participants had limited follow-up coaching. In South Sudan, the implementing NGO's local office shut down in late 2013, about half a year after the productive asset transfers and during follow-up coaching visits. In Yemen, household coaching visits were sometimes canceled due to road closures or unrest, and consumption support was suspended for over a year during the two-year program.

with high fidelity to their original designs.

Income is the most frequently measured outcome, observed in 52 arms and 82 arm-waves. The average effects for income are broadly consistent with those for consumption, with an increase in monthly household income of +28.0 USD (2020 PPP), or 15 percent relative to the median control mean of 190 USD (2020 PPP).[14] Figure 2 shows that the treatment arms that generated the largest income effects are mostly the same as those that generated the largest consumption effects: the Graduating to Resilience Program in Uganda (Beam et al., 2026), the WINGS program in Uganda (Blattman et al., 2016), and the Afghanistan Targeting the Ultra Poor program (Bedoya et al., 2019). In addition, the "full" versions of the Bab Amal program in the Arab Republic of Egypt (Assaad et al., 2026) and of the government-led adaptive social protection programs in Niger (Bossuroy et al., 2022) and Senegal (Bossuroy et al., 2025b) delivered large income effects. Conversely, "full" interventions in Burkina Faso (Bossuroy et al., 2024), and streamlined interventions in India (Bauchet et al., 2015) and Uganda (Sedlmayr et al. (2020), microenterprise program) generated non-significant improvements in income. In Ethiopia, the initial evaluation found positive effects on income and revenues (Banerjee et al., 2015a), but these effects were smaller and no longer statistically significant at the seven-year follow-up (Barker et al., 2024). In a highly insecure area of Burkina Faso, the program faced implementation issues, but participants also saved more than they invested. On the other hand, in Ethiopia and India, there seem to have been promising economic opportunities for control group households during the study periods, potentially resulting in control households' income catching up with treatment households.

The point estimates for consumption and income are close (+38.1 and +28.0 in 2020 PPP USD, respectively), with overlapping credible intervals.[15] The median control mean for consumption is much larger than that for income (361 and 190 in 2020 PPP USD, respectively). We note, however, that the measurement of consumption tends to be more consistent across studies than income. Specifically, not all studies report total household income considering all income sources (agriculture, livestock, off-farm businesses, wage employment, and others), as well as for all individuals in the households. We take the measure closest to total income we can find in each study, but the fact that some studies only report partial income implies that we might underestimate total income and possibly also treatment effects.

Food security is measured in slightly fewer arms (41 arms; 61 arm-waves). Results on food security also show statistically significant and substantial impacts of 0.20 standard deviation, confirming that Graduation programs achieve broad gains in economic well-being. There is some overlap in the programs that delivered relatively large or small impacts on food security and consumption.[16] Yamba Malawi's program, which emphasizes children's nutrition and food security (Poll, 2025), the Graduating to Resilience Program in Uganda (Beam et al., 2026), World Vision's program in Somalia (Leight et al., 2025), and the Niger adaptive safety net program (Bossuroy et al., 2022) had relatively large

[14] See Appendix Table A.3 for the estimates.

[15] Appendix Figure A.3 shows that income and consumption have a Spearman rank correlation of 0.32.

[16] See Appendix Figure A.2, Panel C. The food security and consumption treatment effects have a rank correlation of 0.41, see Appendix Figure A.3.

impacts on food security. In contrast, results from "full" programs in multiple countries are small or imprecise enough that many do not achieve statistical significance (Hirvonen et al. (2025), Bossuroy et al. (2024)), including in urban Senegal where food insecurity might be less of a concern (Bossuroy et al. 2025b). The Swayam Krishi Sangam program in India (Bauchet et al., 2015) did not generate improvements in consumption, income, or food security.

Savings is measured less frequently (25 arms; 30 arm-waves). Appendix Figure A.2, Panel D, shows that impacts on savings tend to be positive, with an increase in amounts saved of +55 USD (2020 PPP). This represents more than 10 times the median control mean; the 95 percent credible interval ranges from 5 to 104 USD (2020 PPP). The results on savings are somewhat noisy, however, and influenced by the very large effects found in Ethiopia by Banerjee et al. (2015a), where participants were required to regularly deposit savings over the course of the program, and were unable to withdraw funds until they reached a savings balance equal to the value of the asset transfer. Excluding this estimate, the pooled effect remains positive but falls to +30.2 USD in 2020 PPP (95 percent credible interval: 14.2 to 46.0 USD in 2020 PPP). Appendix Table A.5 reports how excluding the Ethiopia estimate changes the pooled savings effect. In general, interpreting these effects also warrants caution, because savings is not measured fully consistently across studies. First, studies differ in whether they capture a stock or flow of savings. Second, studies differ in whether they measure total savings or savings held in a specific instrument, and few capture substitution across instruments. Because several programs actively promote a particular savings vehicle, a measured increase in savings within that vehicle need not reflect higher household savings overall. We take the most comprehensive measure of total savings available in each study, but these differences imply the pooled savings estimate is noisier and less comparable across studies than consumption. In addition to the Ethiopia pilot, the WINGS program in Uganda (Blattman et al., 2016), group coaching variants in Uganda and Bangladesh (Beam et al., 2026), the Afghanistan Targeting the Ultra Poor program (Bedoya et al., 2019), and BRAC's hybrid credit model in Bangladesh (Rahman et al., 2021) all have large positive impacts on savings. In both Uganda programs, participants received training on establishing savings groups; in Afghanistan and both Bangladesh programs, coaches encouraged participants to save. The treatment arms with the smallest impacts on savings were a micro-franchising intervention in Kenya (Brudevold-Newman et al., 2024) and a streamlined intervention in Nepal that had overall muted impacts (Janzen et al., 2021).

Assets are the third-most reported outcome, with estimates from 29 studies, 40 arms, and 60 arm-waves. The estimated effect on total assets is +282 USD in 2020 PPP, 40 percent of the median control mean. As for income, we take the most comprehensive measure available in each study, though their scope may vary; in particular, some interventions provide in-kind transfers made of specific assets, while others offer cash grants that may be invested in a broader range of assets, saved, or consumed. The programs that provided the largest asset transfers generated some of the largest impacts on total assets, a pattern we will discuss further below (in Section 4.2). For both savings and assets, effects seem highly correlated with the value of the asset transfer provided as part of the program.[17]

[17] These outcomes are also highly rank-correlated with each other (0.48, see Appendix Figure A.3).

Overall, results on consumption, income, savings and assets indicate broad-based positive impacts on economic well-being, showing that Graduation programs induce robust improvements in economic outcomes for the low-income households they target.

4.1.2 Employment

Graduation programs can induce a range of potential occupational shifts. Some interventions focus on supporting agricultural or livestock activities, while others aim to boost off-farm businesses and diversification. Many studies document impacts on a range of income-generating activities, yet the available studies are not always fully consistent in measuring employment across various types of occupations.

Panel F of Appendix Figure A.2 presents results for the variable that is closest to capturing "any type of work". The posterior average effect is positive but small and imprecisely estimated at 0.04 (95 percent credible interval: 0.01 to 0.09). On the one hand, finding only a small impact on any form of employment may not be surprising in high-poverty contexts where formal unemployment is low. On the other hand, and for the same reason, the control mean of this measure is 33%, suggesting that it may not appropriately capture all forms of employment, including in agriculture and livestock activities as well as wage work and self-employment.

Appendix Figure A.2, Panel G, provides results for the probability of engaging in off-farm self-employment, i.e. the operation of household businesses or microenterprises. This is a narrower variable that the one capturing any employment, and not all Graduation programs aim to promote off-farm businesses, but it has the advantage of being more consistently measured. Results show a positive and significant impact on business ownership of 8 percentage points, or 27 percent relative to a control mean of 0.29. The increase in business ownership is particularly strong among participants in Uganda's WINGS program (Blattman et al., 2016), with large gains also observed in Yamba Malawi's program (Poll, 2025).

4.1.3 Psychological Well-Being and Agency

Panels H, I and J of Appendix Figure A.2 present results for dimensions of psychological and social well-being: mental health (measured in 22 studies, 33 arms, and 49 arm-waves), as well as life satisfaction and aspirations (less frequently observed). The exact measures vary between studies, but we group them by domain and present results for standardized indices.

Results show consistent improvements across these various indices of psychological well-being. Mental health improves by 0.10 standard deviations, life satisfaction/happiness by 0.20 standard deviations, and aspirations/expectations by 0.09 standard deviations. Our results suggest mental health effects of similar magnitude to those documented in Romero et al. (2021), who include eight of the 81 treatment variations in our sample in their meta-analysis on the psychological effects of economic transfers. They

find that Graduation programs generate a 0.08 standard deviation improvement in mental health, with unconditional cash transfers generating an average 0.15 standard deviation improvement, and asset transfers generating an average 0.16 standard deviation improvement. Meta-analyses focused on unconditional cash transfers alone find similar effects on mental health: McGuire et al. (2022) find an effect of 0.10 standard deviations, and Crosta et al. (2026) find an effect of 0.15 standard deviations. Overall, our results indicate that Graduation programs do not just improve economic well-being, but also induce broad gains in psychological well-being and life satisfaction consistent with those generated by other transfer programs.

As seen in Panel H of Appendix Figure A.2, most treatment arms that measure mental health find positive effects, though some do not achieve statistical significance. The Afghanistan Targeting the Ultra Poor program (Bedoya et al., 2019), the full adaptive safety net program in Niger including a cash grant and psychosocial components (Bossuroy et al., 2022), and the Graduating to Resilience program in Uganda (Beam et al., 2026) generate particularly large effects, possibly reflecting the high-touch nature of the support provided. None of the programs at the bottom of the distribution stand out relative to the rest of the treatment arms.[18]

Beyond psychological well-being, Graduation programs also have positive impacts on women's empowerment and agency (see Panels K and L of Appendix Figure A.2). Agency is measured in 23 arms (32 arm-waves), while women's empowerment is measured in 13 arms (22 arm-waves). The specific measures again vary between studies, and we document effects on standardized indices. On average, Graduation and economic inclusion programs improve women's agency by 0.09 standard deviations, significant at the 5% level. While the initial set of Graduation pilots in Banerjee et al. (2015a) had limited effects on agency and women's empowerment, the broader set of studies covered in the meta-analysis shows positive effects. Women's empowerment increases by 0.08 standard deviations.

Panel K of Appendix Figure A.2 shows that the treatment arms with the largest impacts on women's agency include all three adaptive safety net intervention modalities in Niger (Bossuroy et al., 2022), the full variation in Mauritania (Bossuroy et al., 2025a) and a streamlined intervention in Nepal (Janzen et al., 2021), all of which explicitly targeted women as the primary recipient of support. Point estimates from Burkina Faso (Bossuroy et al., 2024) are slightly negative though not significantly so in a context where cash grants were delivered to younger women within households.

4.1.4 Correlations across Outcomes

A natural question is whether the impacts between outcomes move together: do programs that improve one outcome tend to improve others, or are programs heterogeneous in the types of effects they generate? Appendix Figure A.3 summarizes the cross-outcome Spearman rank correlations of treatment effects across arms. Two broad patterns emerge. First, economic outcomes appear to cluster together: as mentioned above, effects on total assets and amount saved are highly rank correlated (0.48), con-

[18]The pattern is broadly similar for life satisfaction and aspirations, but we do not discuss the results in detail due to the number of observations being more limited.

sistent with both being potentially associated with the value of the transfers. Savings effects are also rank-correlated with household monthly consumption (0.53), while the corresponding association for total assets is weaker (0.27). Second, treatment effects on food security are strongly correlated with those on psychological and social well-being, most strongly with the life satisfaction and happiness index, suggesting that improvements in day-to-day food security and in how people feel about their lives are associated.

4.1.5 Heterogeneity across Treatment Arms and follow-up waves

We compare baseline and component-model heterogeneity on the same complete-case treatment-arm–follow-up sample for each outcome (Appendix Table A.6). All eligible follow-up waves are retained. The heterogeneity standard deviation captures variation in the underlying, or "true" treatment effects after accounting for the sampling uncertainty in each reported estimate. This dispersion is of the same order of magnitude as the mean effect for every outcome and exceeds it for several outcomes. For comparison, the unrestricted pooled mean monthly consumption effect is 38.1 USD in 2020 PPP, while the matched-sample baseline heterogeneity standard deviation is 37.1; for total assets, the corresponding estimates are 282.4 and 445.6. Therefore, the variation in observed estimates is not simply a consequence of some studies being estimated imprecisely: the model uncovers economically meaningful differences in underlying impacts across programs and settings.

The pooling factors reported in Appendix Table A.6 provide a complementary way of describing this heterogeneity. They measure the extent to which variation in treatment-arm estimates is attributed to sampling uncertainty rather than differences in underlying effects, and therefore the degree to which individual estimates are pulled toward the common mean. Pooling factors are below 0.5 for every outcome and below 0.3 for most outcomes, indicating that treatment-arm estimates generally retain more weight than the overall mean when estimating their underlying effects. The particularly low pooling factors for savings, assets and employment imply that differences across arms are large relative to their sampling uncertainty. These low pooling factors underscore an external validity challenge: the average impact alone provides an incomplete guide to the effect likely to arise from any particular treatment arm.

4.2 Impacts by Program Components and Characteristics

We now explore systematically how impacts vary by the composition of the multifaceted Graduation program bundle and other characteristics. The following sections then expand the analysis by implementation year (Section 4.3), time since the end of the program and costs (Section 4.4).[19]

Figure 3 compares treatment effects for programs with all five components against those with only three or four. More comprehensive programs with all five components tend to have larger effects. The difference is most conclusive for assets, which increase for comprehensive programs but not for programs

[19]For simplicity, we refer to "program" in this discussion as a shorthand for "treatment arm".

with 3 or 4 components. Results for other outcomes are less clear-cut: the posterior probability that programs with all five components have larger effects is $Pr(\Delta > 0) = 0.87$ for consumption, $Pr(\Delta > 0) = 0.81$ for income, $Pr(\Delta > 0) = 0.87$ for food security, and $Pr(\Delta > 0) = 0.81$ for savings. Programs with all five components also tend to have larger effects on life satisfaction ($Pr(\Delta > 0) = 0.93$) and (slightly less conclusively) on mental health ($Pr(\Delta > 0) = 0.84$). The data on other outcomes, including employment and empowerment indices, are more limited.

Figure 4 compares treatment effects on consumption and income with different program compositions, namely with and without coaching, savings support, consumption support, and a large transfer (a total value of at least 400 USD in 2020 PPP for the productive transfer plus consumption support). The differences are incremental associations from a simultaneous additive meta-regression. We cannot separately estimate the association with training because every Graduation arm contributing to the consumption and income analyses includes training.[20]

The clearest result on program composition is that transfers of at least 400 PPP USD have a positive association with consumption ($Pr(\Delta > 0) = 0.98$) and income ($Pr(\Delta > 0) = 0.98$).[21] Variance is high, however, and some potentially important outliers demonstrate that a large lump sum transfer is not necessary to generate large treatment effects (e.g., the psychosocial arm without cash grants in Bossuroy et al. (2022) and the similarly large results in Mahmud and Riley (2026)).

Critically, however, Panel B shows that the higher treatment effects for programs with large transfers do *not* lead to higher cost-effectiveness. The estimated differences in monthly treatment effects per US$100 of cost are slightly negative for both outcomes, with wide credible intervals spanning zero. We expand on the discussion on cost and cost-effectiveness in Section 4.4 below.

Besides large transfers, results on the value-added of specific components do not show a consistent pattern whereby any other component stands out as either necessary or irrelevant. The inclusion of broad savings support is associated with smaller impacts on income. However, the most common form of savings support involves savings group, and savings groups do not have a significantly different effect on income ($Pr(\Delta > 0) = 0.91$), suggesting that the negative effect of savings support is driven by the small number of studies without savings groups (see Appendix Table A.9 for point estimates).

The bottom of each panel in Figure 4 documents results for additional characteristics, including whether governments are involved, whether programs are small pilots or not (proxied by the size of the program being larger than the size of the evaluation sample), and whether programs were implemented through 2015 or in 2016 or later. The first two speak to a broader concern that effects measured in initial evaluations often attenuate as programs expand beyond pilots or move into government delivery (Allcott, 2015; Banerjee et al., 2017; Muralidharan and Niehaus, 2017; Bold et al., 2018; Vivalt, 2020).

We first consider scale, proxied by whether program coverage exceeds the size of the evaluation sample. We find no evidence that larger programs yield smaller treatment effects. If anything, point estimates on consumption ($Pr(\Delta > 0) = 0.98$) and income ($Pr(\Delta > 0) = 0.57$) are larger. The proxy

[20] See Appendix Figure A.4 for plots of all outcome variables using three transfer-size categories: below 400, 400–799, and 800 or more in 2020 PPP USD. Appendix Table A.7 reports the corresponding estimates.

[21] Appendix Table A.8 shows that results are robust when including partly repayable grants.

is admittedly coarse and does not distinguish moderately sized from national programs. Still, this suggests that effects do not necessarily dilute as pilots expand or scale.

In contrast, programs with government involvement tend to have lower treatment effects on consumption ($Pr(\Delta > 0) = 0.13$) and income ($Pr(\Delta > 0) = 0.02$). However, they are no different on cost-effectiveness terms (Panel B), which is a by-product of government programs operating at a lower cost, and lower cost programs being more cost-effective, as we discuss further in Section 4.4. Indeed, several government-led programs reach among the highest benefit-cost ratios in our sample (Data Appendix Table 4). And some programs with government involvement in Niger, Zambia and Senegal have demonstrated large impacts at relatively large scale (Bossuroy et al., 2022; Botea et al., 2023; Bossuroy et al., 2025b), again pointing to substantial heterogeneity in the results of government-led programs.

Lastly, in the next section we discuss the final row in Figure 4, implementation year, which finds that more recently implemented programs performed considerably better than earlier programs.

4.3 Impact by Implementation Year

Treatment effects have increased over time in the past 20 years. We see this as a simple binary result in Figure 4: in the bottom row, programs implemented after 2015 generate +39 PPP USD per month more in consumption ($Pr(\Delta > 0) = 0.98$) and +36 PPP USD in income ($Pr(\Delta > 0) = 1.00$).[22] This is more than the effect of any one component shown in Figure 4. The point estimates favor more recent programs both for average treatment effects irrespective of costs (Panel A) and for effects per US$100 PPP of costs (Panel B), although the cost-normalized contrasts are less precise and their 95 percent credible intervals include zero. These results emerge even though government involvement became more common over the same period.

Turning to a more granular examination, Figure 5 presents a scatter plot and linear regression line, which shows that every five years is associated with an additional 9.1 USD PPP in monthly income effects and 7.1 USD PPP in monthly consumption effects.

These patterns could mask many confounds. Costs having also gone down, not up, this is not a by-product of more intense programs with respect to resources. In fact, we find the opposite: average treatment effects per $100 of cost have a similarly upward trend (row 2 figures in Figure 5). Appendix Table A.10 reports value-for-money trends for the main outcomes. Not only have average treatment effects risen, but the efficiency with which the resources have been deployed has also improved. We also then examine results after partialing out costs (row 3 figures in Figure 5). Understanding this pattern is not straightforward, and we explore several potential explanations.

First, we assess whether the publication process is responsible for more recent programs having larger effects. The earlier studies in our sample are more likely to have completed peer review, and if refereeing shrinks reported effects–through more conservative inference or the attrition of favorable

[22] Although the median implementation year across the 68 unique arms is 2017, we observe a clear break with no program starting implementation in 2015. Splitting the sample at 2015, we have 23 arms in the earlier group and 45 in the later group.

specifications–then a study's age would be mechanically associated with a smaller reported effect. To test this hypothesis, we consider the unconditional cash transfer literature over the same period as a placebo, drawing on Crosta et al. (2026). Results are shown in Figure 6, and Appendix Figure A.6 brings together the Graduation, cash-transfer, and financial-literacy comparisons. Importantly, we find a flat trend over time for cash, which, if anything, is mildly downward sloping.[23] The lack of similar upward trend suggests that the publication bias, composition of working paper versus published articles, changes in the process of conducting RCTs or measuring outcomes, or timing of release of papers, etc., is not the underlying cause of the observed upward trend over time for the Graduation studies.

Second, we consider whether more recent programs are simply better designed, perhaps building on earlier evidence. If later programs select a better mix of components, adjusting for program design should absorb part of the trend. The fourth row of Figure 5 plots treatment effects adjusted to the sample means of large transfers, consumption support, coaching, savings support, and organized savings groups within each posterior draw.[24] The positive trend by implementation year remains after this adjustment, indicating that changes in observed component composition do not explain the pattern.[25] These results align with Section 4.2 above, where no bundle comes out reliably ahead aside from large transfers. There is no obvious lesson from the pre-2016 trials suggesting that there might be an optimal bundle either.

Third, we examine whether programs became more efficient over time. Early enthusiasm about the effects of Graduation programs was paired with persistent concern in the broader policy community about their cost, and many implementers and government practitioners sought models that could deliver the same results for less. But with average treatment effects increasing over time, not just treatment effects per dollar cost, added costs would need to actually lower treatment effects for this to be the case.[26] The third row of Figure 5 plots treatment effects adjusted to the sample mean of predicted program cost within each posterior draw. The positive implementation-year trend remains after this adjustment, indicating that changes in program cost do not explain the pattern.

Fourth, we turn to the possibility that practitioners in the Graduation community improved operational efficiency, which could include design features but not ones captured by the broad and perhaps overly simplistic binary categorization of components. These could include even more difficult to parameterize implementation details such as content of staff training and hiring practices, implementation quality, monitoring and feedback loops to field agents,... The broader "community" of Graduation implementers and donors engaged in frequent events to exchange information, engage in mutual site visits, etc., and such learning could be a by-product of that community effort. We (currently) are unable to test this hypothesis causally, but note that, over the period our sample covers, Gradua-

[23] Readers should not compare individual point estimates from Graduation in Figure 5 and cash in Figure 6, as cash is scaled by the amount transferred while Graduation is either not scaled or scaled by total program costs.

[24] See the adjusted comparison in Appendix Table A.9.

[25] Appendix Figure A.7 extends the component-adjusted specification to all main outcomes.

[26] In theory this could be true if layers of complicated programs lower the focus of interaction with recipients. While this may be true for a few cases, we argue this is unlikely to be the root cause of the overall trend.

tion implementers made sustained efforts for transmitting details on the practice. This included the production of toolkits and manuals by BRAC, other NGOs and early government adopters. It also included regular exchanges between Graduation practitioners, which were facilitated by CGAP and the Partnership for Economic Inclusion (PEI), annual researcher and implementer gatherings organized by Innovations for Poverty Action and its partners, as well as regional efforts such as communities of practices on Social Protection or the Sahel Adaptive Social Protection program (see for instance some of the examples and case studies discussed Andrews et al. (2021)). These activities supported learning between practitioners and facilitated government adoption.

Lastly, we also examine the trend by implementation year of financial literacy education on financial behavior, using the published meta-analysis and publicly available effect-size data and study characteristics from Kaiser et al. (2022b). We find a similarly upward trend (Row 3 of Figure 6). Row 4 partials out the features identified in that meta-analysis: instruction intensity, participant age group, low-income targeting, country-income group, intervention channel, financial-behavior outcome domain, publication in a top economics journal, and follow-up timing. The annual trend is nearly unchanged.[27] Thus changes in observed program features do not account for the trend. While we are not attempting to compare the knowledge exchange across these two groups of practitioners, we note that for both these types of intervention the implementation quality has more degrees of freedom than for cash, and thus the upward trend pattern is consistent with a community which learned and iterated in a way that could viably improve operational efficiency. While for cash some design and delivery features (e.g., timing, magnitude, messaging) may influence relative effectiveness, we assert that the lever of operational efficiency is likely less for cash than for the Graduation and financial education programs.

In line with the above results, the learning was not about finding an optimal component bundle or streamlining costs. Instead, improved program effectiveness may have emerged through features we do not observe consistently across studies, e.g., delivery and implementation modalities, the content and intensity of training and coaching, the quality of frontline delivery and supervision,... Appendix Table A.6 demonstrates that high-level program design explains a relatively small share of program impacts, leaving a strong potential role for such factors.[28]

[27] Specifically, it goes from 0.0097 Hedges' g (95% CrI [0.0031, 0.0164]; Pr(slope $>$ 0) = 99.8%) to 0.0101 (95% CrI [0.0031, 0.0171]; Pr(slope $>$ 0) = 99.7%).

[28] Indeed, a large part of the heterogeneity in impacts remains unexplained by the observable program characteristics included in our model. The variance decomposition in Appendix Table A.6 separates variation associated with coaching, savings support, consumption support, and transfer size from residual variation across treatment arms and follow-up waves. For eleven of the twelve outcomes, these characteristics account for approximately 6–26 percent of modeled heterogeneity. The estimated share is substantially higher only for empowerment, at 93 percent, although the corresponding credible interval is wide. Substantial residual heterogeneity may reflect differences in economic context, target populations, implementation, and other unobserved dimensions not captured by the model.

4.4 Persistence and Cost-Benefit Analysis

4.4.1 Persistence of Impacts

Above we documented robust and broad-based impacts of Graduation programs on economic outcomes, as well as on a range of well-being and empowerment dimensions. At the same time, the estimates indicated substantial variation in the impacts observed between programs. In this section, we examine the persistence of impacts across follow-up surveys. In doing so, we focus on the main economic outcomes, consumption and income, which are the most frequently measured in the literature.

Most interventions in our sample last between 12 and 24 months. In parallel, the majority of follow-up surveys document impacts between 6 months and 48 months after the end of the program, although there are now a few studies with follow-ups after 6 years or more in Bangladesh, Ethiopia, and India. Figure 7 examines the persistence of consumption and income treatment effects across follow-up surveys.

Figure 7 (Panel A) suggests that consumption effects strengthen at later follow-ups. The estimated annual slope is 47.8 USD PPP, and 94.2 percent of the posterior distribution is positive, although the 95 percent credible interval remains wide (−11.5 to 107.5). For income (Panel B), the estimated change in the annual household effect per follow-up year is 16.0 USD PPP, with a 95 percent credible interval of −35.9 to 68.9 and a 73.0 percent posterior probability of a positive slope. The point estimate favors increasing effects, but the interval also allows flat or declining effects. Restricting the persistence regression to follow-ups one to four years after baseline yields a similar central consumption trajectory, although uncertainty is wider (Appendix Figure A.8). The income trajectory is more sensitive to this restriction, with a negative central slope and a credible interval that includes zero. Appendix Figure A.9 shows substantial heterogeneity in within-arm changes across follow-up waves.

Taken together, the evidence favors increasing consumption effects, while the income trajectory is estimated less precisely and meaningful increases or declines cannot be ruled out.[29]

4.4.2 Impacts by Cost and Cost-Benefit Analysis

Figure 8, row 1, plots annual household consumption and income treatment effects against predicted program cost per household.[30] For this analytical comparison, costs and treatment effects are both expressed in constant 2020 PPP USD. Treatment effects are taken from the observed follow-up surveys, without assuming any persistence. Results for consumption show that more expensive programs have larger treatment effects. This is in line with the results above showing that programs with larger transfers have larger treatment effects. For income, however, treatment effects do not vary much by predicted costs, though the credibility interval is again broader for income than for consumption.

A similar analysis can be performed by plotting treatment effects per unit of (predicted) cost.

[29] Figure A.10 shows results for a larger set of outcomes, though the number of studies and variation in follow-up timing are more limited than for consumption and income. Qualitatively, results for assets and savings point to sustained gains. Some effects on mental health and life satisfaction appear to fade, but these patterns are based on fewer observations.

[30] The horizontal axis uses predictions from the full cost model, which combines available cost metadata and program components, as discussed in Section 2.4.

Figure 8, row 2, shows that treatment effects on consumption and income per $100 of predicted cost tend to decline as costs rise. This suggests that treatment effects per dollar are larger for relatively less expensive programs. To put this another way, treatment effects increase with costs, but at a lower rate; hence the more cost-effective programs tend to be lower cost. Row 3 reports descriptive value-added relative to the fitted cost benchmark, i.e. each treatment-effect from row 1 minus the fitted effect at the same predicted cost. Positive values indicate effects above the benchmark and negative values effects below it.[31]

In order to perform a more careful cost-effectiveness analysis, we take into account the persistence of impacts. Figure 9 presents benefit-cost ratios for programs with reported cost data. The calculations include consumption benefits only.[32] We report three persistence scenarios. First, a conservative "linear-decay" scenario assumes that benefits decay linearly over 8 years, similar to the patterns observed in Ethiopia by Barker et al. (2024). Second, an "estimated persistence" scenario is based on the persistence slope estimated from our sample: it projects benefits over a common 10-year horizon integrating over the full posterior distribution of the consumption persistence slope and flooring projected effects at zero.[33] Third, a "permanent persistence" scenario follows the cost-benefit analysis estimation in Banerjee et al. (2015b) and assumes that the final observed annual consumption effect persists forever. For all three cases, we construct each benefit-cost ratio from latent effect draws, preserving hierarchical shrinkage and effect uncertainty. We can use 43 treatment arms from 32 study records. After grouping multiple arms and repeat evaluations of the same intervention lineage, they represent 31 underlying programs, which are the units given equal weight in the pooled calculation. We also provide subgroup comparisons by program composition in Appendix Table A.11, noting that these are descriptive rather than causal.

The first row of Figure 9 provides the overall benefit-cost ratio: 2.51 under linear decay, 4.34 under estimated persistence, and 9.55 under permanent persistence. The estimated persistence scenario is our preferred specification because it uses the pattern observed in our data, although the number of studies with long-term follow-ups of course remains limited. The linear decay scenario and the permanent persistence scenario provide alternative benchmarks for the benefit-cost ratios. All three estimates show that Graduation programs are highly cost-effective, with benefits 2.51–9.55 times larger than costs. Benefits are 4.34 times costs under the estimated-persistence scenario.[34]

Given our results that treatment effects are substantially larger for Graduation programs implemented more recently, we also calculate benefit-cost ratios separately for programs implemented through 2015 and those implemented in 2016 or later. Panel B of Figure 9 shows substantially higher

[31] Appendix Table A.17 compares these relationships using reported and predicted PPP costs, both on their available samples and on a common sample.

[32] Adding productive assets to consumption benefits risks counting transferred assets once as a stock and again through the consumption they generate.

[33] The selection of a 10-year horizon was informed by the fact that it is the longest follow-up data in our sample (Banerjee et al., 2021). This is not meant to imply that effects will revert to zero after 10 years, however.

[34] To assess sensitivity to selective long-term follow-up, we re-estimate consumption persistence using only follow-ups one to four years after baseline. Holding the treatment-effect anchors, program costs, discounting, and 10-year benefit horizon unchanged, the benefit-cost ratio is 4.50 (95 percent credible interval: 1.85–8.01), compared with 4.34 under the full-sample persistence estimate. Appendix Figure A.8 shows similar central consumption trajectories, with wider uncertainty under the restricted specification.

benefit-cost ratios for recent programs. Under estimated persistence, the ratio is 5.39 [3.80, 7.00] for recent programs and 2.89 [1.57, 4.27] for earlier programs. This corresponds to increases of approximately 87 percent. Appendix Table A.12 reports the underlying estimates for all three persistence scenarios.[35]

Panel B of Figure 9 compares benefit-cost ratios for unconditional cash transfers (obtained from Crosta et al. (2026)) under the same three persistence assumptions.[36] Among programs implemented through 2015, cash has a higher ratio under linear decay (2.03 versus 1.36) and permanent persistence (9.68 versus 4.35), while Graduation is slightly higher under estimated persistence (2.89 versus 2.50). Among programs implemented in 2016 or later, Graduation has the higher ratio under linear decay (3.34 versus 1.61), estimated persistence (5.39 versus 1.93), and permanent persistence (13.32 versus 7.95).

5 Conclusion

Leveraging a rapid expansion of Graduation programs and related research, our meta-analysis shows that Graduation programs generate large, durable improvements in the economic well-being of low-income households. Average impacts on consumption (+38.1 USD PPP, 11 percent) and income (+28.0 USD PPP, 15 percent) are accompanied by gains in savings, assets, food security, mental health, and women's empowerment. Importantly, gains in consumption and income sustain for several years after program completion.

Graduation programs are multifaceted by nature, and we analyze results by program composition. Programs with transfers of at least 400 USD PPP are associated with larger impact on consumption and income, although the increase is commensurate with the additional cost of the transfer (thus on cost-effectiveness terms, programs perform similarly with or without a large transfer). More broadly, treatment effects increase with costs, but at a lower rate; hence the more cost-effective programs tend to be lower cost. Beyond the inclusion of large transfers, other components do not tell a clear story, and no component is rendered either essential or unnecessary. Considerable heterogeneity persists even after accounting for observable design features.

Perhaps the most striking finding is that programs implemented after 2015 generate substantially larger treatment effects on consumption and income than earlier cohorts, with benefit-cost ratios growing from 2.89 for earlier programs to 5.39 for recent programs under our preferred estimated-persistence scenario, a 87 percent increase.

This improvement cannot be explained by publication timing, by shifts toward a more optimal

[35] Under linear decay, the ratio is 3.34 [2.85, 3.83] for recent programs, compared with 1.36 [0.85, 1.88] for earlier programs. Under permanent persistence, it is 13.32 [11.24, 15.41] for recent programs and 4.35 [2.49, 6.20] for earlier programs. These correspond to increases of approximately 145 and 206 percent, respectively.

[36] For estimated persistence, cash benefits are projected over a 10-year horizon, integrating over the full posterior distribution of the pooled cash consumption follow-up slope and flooring effects at zero. Under permanent persistence, each program's final observed annual effect is held constant forever. The Graduation denominator remains the full reported program cost in constant 2020 PPP USD. The cash denominator is the amount transferred plus an administrative-cost allowance equal to 24 percent of the transfer, following the benchmark reported in Table 11 of Crosta et al. (2026).

component bundle, or by mere cost reductions. At the same time, observable program and study characteristics explain only a modest share of the heterogeneity in impacts across treatment arms, suggesting that much of what separates more and less effective programs still lies in implementation details we do not consistently observe or measure. The evolution of program effectiveness over time could thus be partly explained by Graduation practitioners learning how to better deliver programs through knowledge exchanges and shared tools, which the field has made dedicated efforts to nurture over the past decade. This suggests that continued investment in cross-program learning together with context-specific testing is likely to remain a productive path for improving the effectiveness of economic inclusion programs going forward.

We posit that development economics would benefit from greater attention to implementation science. If, as our results suggest, the devil is in the details of implementation, there may be large returns to identifying which of those details matter and how high-quality implementation can be generated and replicated. By "implementation quality", we mean not only fidelity to a specified program design, but also the more granular choices that determine how that design is delivered—features far more nuanced than whether a program, for example, includes "coaching". A growing literature illustrates the importance of these distinctions. Seemingly similar interventions can generate quite different impacts depending on how faithfully they are implemented, who delivers them, and the specific arrangements through which they are operationalized (Banerjee et al., 2017; Angrist and Meager, 2023; Bandiera et al., 2023). Our meta-analysis findings suggest value in measuring implementation with the same care with which development economists increasingly measure treatment effects, and similarly that experimental designs should focus on implementation decisions and not just high-level design decisions.

More broadly, these results highlight an agenda for development economics that looks beyond specific interventions and policies (regardless of how well-grounded the inquiry is in theory) to delve more into how programs themselves learn, iterate, and share, and what policy domains have more room for such efforts to yield productivity gains. Such inquiries are akin to a large and old literature from business and industry, one that emphasizes knowledge diffusion and spillovers as important drivers of productivity growth (Bloom et al., 2013; Crespi et al., 2008). Applying these ideas to development interventions raises a set of questions beyond the scope of this paper: Do some policy practitioner communities improve more rapidly than others, and why? When they do, is improvement driven by diffusion of knowledge or merely by each individual organization improving within its own setting, data, and operations? Are there "system-change" level investments that global donors can make that improve the diffusion of knowledge? Are there global lessons about intervention design, implementation and operational delivery, and generalizable insights about the internal learning process in which each entity needs to engage in order to improve? Systematically comparing changes in effectiveness over time across policy interventions could help answer these questions, alongside attempts to document more precisely how design and operating processes change within organizations.

References

Al-Ubaydli, O., List, J. A., and Suskind, D. L. What can we learn from experiments? understanding the threats to the scalability of experimental results. *American Economic Review*, 107(5):282–286, 2017. doi: 10.1257/aer.p20171115.

Allcott, H. Site selection bias in program evaluation. *The Quarterly Journal of Economics*, 130(3): 1117–1165, 2015. doi: 10.1093/qje/qjv015.

Andrews, C., de Montesquiou, A., Arévalo Sánchez, I., Dutta, P. V., Paul, B. V., Samaranayake, S., Heisey, J., Clay, T., and Chaudhary, S. *The State of Economic Inclusion Report 2021: The Potential to Scale*. World Bank, Washington, DC, 2021. doi: 10.1596/978-1-4648-1598-0. URL http://hdl.handle.net/10986/34917.

Angrist, N. and Meager, R. Implementation matters: Generalizing treatment effects in education. EdWorkingPaper 23-802, Annenberg Institute at Brown University, 2023.

Arévalo-Sánchez, I., Heisey, J., Chaudhary, S., Clay, T., Strokova, V., Vasudeva Dutta, P., and Andrews, C. *The State of Economic Inclusion Report 2024: Pathways to Scale*. World Bank, Washington, DC, 2024. doi: 10.1596/978-1-4648-2076-2. URL http://hdl.handle.net/10986/42408.

Assaad, R., Osman, A., Parienté, W., and Valente, C. How big does a "big push" need to be? evidence from randomizing asset transfer size. Working paper, April 24, 2026, 2026.

Bandiera, O., Burgess, R., Das, N., Gulesci, S., Rasul, I., and Sulaiman, M. Labor markets and poverty in village economies. *The Quarterly Journal of Economics*, 132(2):811–870, 2017. doi: 10.1093/qje/qjx003.

Bandiera, O., Burgess, R., Deserranno, E., Morel, R., Sulaiman, M., and Rasul, I. Social incentives, delivery agents, and the effectiveness of development interventions. *Journal of Political Economy Microeconomics*, 1(1), 2023. doi: 10.1086/722898.

Banerjee, A., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Parienté, W., Shapiro, J., Thuysbaert, B., and Udry, C. A multifaceted program causes lasting progress for the very poor: Evidence from six countries. *Science*, 348(6236):1260799, 2015a. doi: 10.1126/science.1260799.

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., Shotland, M., and Walton, M. From proof of concept to scalable policies: Challenges and solutions, with an application. *Journal of Economic Perspectives*, 31(4):73–102, 2017. doi: 10.1257/jep.31.4.73.

Banerjee, A., Duflo, E., and Sharma, G. Long-term effects of the targeting the ultra poor program. *American Economic Review: Insights*, 3(4):471–486, 2021. doi: 10.1257/aeri.20200667. Originally circulated as NBER Working Paper No. 28074, November 2020.

Banerjee, A. et al. A multifaceted program causes lasting progress for the very poor: evidence from six countries. *Science*, 348:1260799, 2015b.

Barker, N., Karlan, D., Udry, C., and Wright, K. The fading treatment effects of a multifaceted asset-transfer program in ethiopia. *American Economic Review: Insights*, 6(2):277–294, 2024. doi: 10.1257/aeri.20230156. URL https://www.aeaweb.org/articles?id=10.1257/aeri.20230156.

Baseler, T., Gazeaud, J., and Snowberg, E. Spillover effects of transfer programs. Aug. 2026. Working paper.

Bauchet, J., Morduch, J., and Ravi, S. Failure vs. displacement: Why an innovative anti-poverty program showed no net impact in South India. *Journal of Development Economics*, 116:1–16, 2015. doi: 10.1016/j.jdeveco.2015.03.005.

Beam, E. A., Brune, L., Das, N., Dercon, S., Goldberg, N., Haque, R., Karlan, D., Khan, M. N., Parkerson, D., Pople, A., Sawada, Y., Udry, C., and Zizzamia, R. Group versus individual coaching for rural social protection programs: Evidence from uganda, philippines, and bangladesh. *American Economic Review: Insights*, January 2026. Conditionally accepted.

Bedoya, G., Coville, A., Haushofer, J., Isaqzadeh, M., and Shapiro, J. P. No household left behind: Afghanistan targeting the ultra poor impact evaluation. Working Paper 25981, NBER, 2019.

Blattman, C., Green, E. P., Jamison, J., Lehmann, M. C., and Annan, J. The returns to microenterprise support among the ultrapoor: A field experiment in postwar Uganda. *American Economic Journal: Applied Economics*, 8(2):35–64, 2016. doi: 10.1257/app.20150023.

Bloom, N., Schankerman, M., and Van Reenen, J. Identifying technology spillovers and product market rivalry. *Econometrica*, 81(4):1347–1393, 2013. doi: 10.3982/ECTA9466. URL https://doi.org/10.3982/ECTA9466.

Bold, T., Kimenyi, M., Mwabu, G., Ng'ang'a, A., and Sandefur, J. Experimental evidence on scaling up education reforms in Kenya. *Journal of Public Economics*, 168:1–20, 2018. doi: 10.1016/j.jpubeco.2018.08.007.

Bossuroy, T., Goldstein, M., Karimou, B., Karlan, D., Kazianga, H., Parienté, W., Premand, P., Thomas, C., Udry, C., Vaillant, J., and Wright, K. Tackling psychosocial and capital constraints to alleviate poverty. *Nature*, 605:291–297, 2022. doi: 10.1038/s41586-022-04647-8. Originally circulated as World Bank Policy Research Working Paper No. 9562 (2021) under the title "Pathways out of Extreme Poverty: Tackling Psychosocial and Capital Constraints with a Multi-faceted Social Protection Program in Niger".

Bossuroy, T., Karlan, D., Kazianga, H., Parienté, W., Premand, P., Udry, C., Vaillant, J., and Wright, K. Impact des mesures d'inclusion productive du programme Burkin Naong Sa Ya au Burkina Faso

(2019–2020). Documents techniques du ppsas, Banque mondiale, Washington, DC, January 2024. Programme de Protection sociale adaptative au Sahel (PPSAS). Licence CC BY 3.0 IGO.

Bossuroy, T., Karlan, D., Kazianga, H., Parienté, W., Premand, P., Udry, C., Vaillant, J., and Wright, K. Impact des mesures d'inclusion productive du programme Tekavoul au Mauritanie. Documents techniques du ppsas, Banque mondiale, Washington, DC, January 2025a. Programme de Protection sociale adaptative au Sahel (PPSAS). Licence CC BY 3.0 IGO.

Bossuroy, T., Karlan, D., Kazianga, H., Parienté, W., Premand, P., Udry, C., Vaillant, J., and Wright, K. Impact des mesures d'inclusion productive du programme Yokk Koom Koo au Senegal. Documents techniques du ppsas, Banque mondiale, Washington, DC, February 2025b. Programme de Protection sociale adaptative au Sahel (PPSAS). Licence CC BY 3.0 IGO.

Botea, I., Brudevold-Newman, A., Goldstein, M., Low, C., and Roberts, G. Supporting women's livelihoods at scale: Evidence from a nationwide multi-faceted program. Working Paper 31625, National Bureau of Economic Research, 2023. URL https://www.nber.org/papers/w31625.

Brudevold-Newman, A., Honorati, M., Ipapa, G., Jakiela, P., and Ozier, O. A firm of one's own: Experimental evidence on credit constraints and occupational choice. *The Review of Economics and Statistics*, 2024. doi: 10.1162/rest_a_01529.

Brune, L., Karlan, D., Kurdi, S., and Udry, C. Social protection amidst social upheaval: Examining the impact of a multi-faceted program for ultra-poor households in yemen. *Journal of Development Economics*, 155:102780, 2022. ISSN 0304-3878. doi: 10.1016/j.jdeveco.2021.102780.

Chowdhury, R., Collins, E., Ligon, E., and Munshi, K. Valuing assets provided to low-income households in South Sudan, Dec. 2016. URL https://economics.ucr.edu/pacdev/pacdev-papers/valuing_assets_provided.pdf. Working paper dated December 15, 2016.

Crespi, G., Criscuolo, C., Haskel, J., and Slaughter, M. Productivity growth, knowledge flows, and spillovers. Working Paper 13959, National Bureau of Economic Research, May 2008. URL https://www.nber.org/papers/w13959.

Crosta, T., Karlan, D., Ong, F., Rüschenpöhler, J., and Udry, C. R. Unconditional cash transfers: A bayesian meta-analysis of randomized evaluations in low and middle income countries. NBER Working Paper 32779, National Bureau of Economic Research, August 2026. URL https://doi.org/10.3386/w32779.

Gobin, V. J., Santos, P., and Toth, R. Poverty graduation with cash transfers: A randomized evaluation. Monash Economics Working Paper 23-16, Monash University, Department of Economics, 2016.

Hirvonen, K., Leight, J., Gilligan, D. O., Mesfin, H. M., Mulford, M., and Tesfaye, H. The impact of a nutrition-sensitive graduation model program on child nutrition: Experimental evidence from

Ethiopia. IFPRI Discussion Paper 02391, International Food Policy Research Institute, December 2025, 2025. URL https://www.jessicaleight.com/uploads/1/3/2/3/13234647/impel_dp.pdf.

Janzen, S. A., Magnan, N., Sharma, S., and Thompson, W. M. Pay it forward: Impacts of a rural livelihoods program with built-in spillovers. (312711), 2021. URL https://ideas.repec.org//p/ags/aaea21/312711.html.

Kaiser, T., Lusardi, A., Menkhoff, L., and Urban, C. Financial education affects financial knowledge and downstream behaviors. *Journal of Financial Economics*, 145(2), 2022a.

Kaiser, T., Lusardi, A., Menkhoff, L., and Urban, C. Financial education affects financial knowledge and downstream behaviors. *Journal of Financial Economics*, 145(2):255–272, 2022b. doi: 10.1016/j.jfineco.2021.09.022. URL https://doi.org/10.1016/j.jfineco.2021.09.022.

Kondylis, F. and Loeser, J. Intervention size and persistence. Technical Report 9769, World Bank, Washington, DC, September 2021. URL http://hdl.handle.net/10986/36242.

Leight, J., Hirvonen, K., and Zafar, S. The effectiveness of cash and cash plus interventions on livelihoods outcomes: Evidence from a systematic review and meta-analysis. Technical Report 2262, International Food Policy Research Institute (IFPRI), Washington, DC, July 2024. URL https://hdl.handle.net/10568/148881.

Leight, J., Hirvonen, K., Karachiwalla, N., and Rakshit, D. Displacement and development: Evidence from a graduation program for Somalia's ultra-poor. IFPRI Discussion Paper 02356, International Food Policy Research Institute, September 2025, 2025. URL https://hdl.handle.net/10568/176389.

List, J. A. *The Voltage Effect: How to Make Good Ideas Great and Great Ideas Scale*. Currency, New York, 2022.

Mahmud, M. and Riley, E. "we're all in this together": Experimental evidence from a universal livelihoods program. Working paper, April 6, 2026, 2026.

McGuire, J., Kaiser, C., and Bach-Mortensen, A. A systematic review and meta-analysis of the impact of cash transfers on subjective well-being and mental health in low- and middle-income countries. *Nature Human Behaviour*, (6):359–370, January 2022. doi: 10.1038/s41562-021-01252-z. URL https://doi.org/10.1038/s41562-021-01252-z.

Meager, R. Understanding the Average Impact of Microcredit Expansions: A Bayesian Hierarchical Analysis of Seven Randomized Experiments. 11(1):57–91, 2019. ISSN 1945-7782, 1945-7790. doi: 10.1257/app.20170299. URL https://pubs.aeaweb.org/doi/10.1257/app.20170299.

Muralidharan, K. and Niehaus, P. Experimentation at scale. *Journal of Economic Perspectives*, 31 (4):103–124, 2017. doi: 10.1257/jep.31.4.103.

Poll, M. Poverty alleviation at high frequency. Job market paper, Brown University, November 26, 2025, 2025. URL https://www.moritz-poll.com/.

Rahman, A., Bhattacharjee, A., and Das, N. A good mix against ultra-poverty? evidence from a randomized controlled trial (RCT) in Bangladesh. *Review of Development Economics*, 2021. doi: 10.1111/rode.12809. Source file provided as local filename "270 Rehman et al A good mix against ultra-poverty.pdf"; online link not available in spreadsheet.

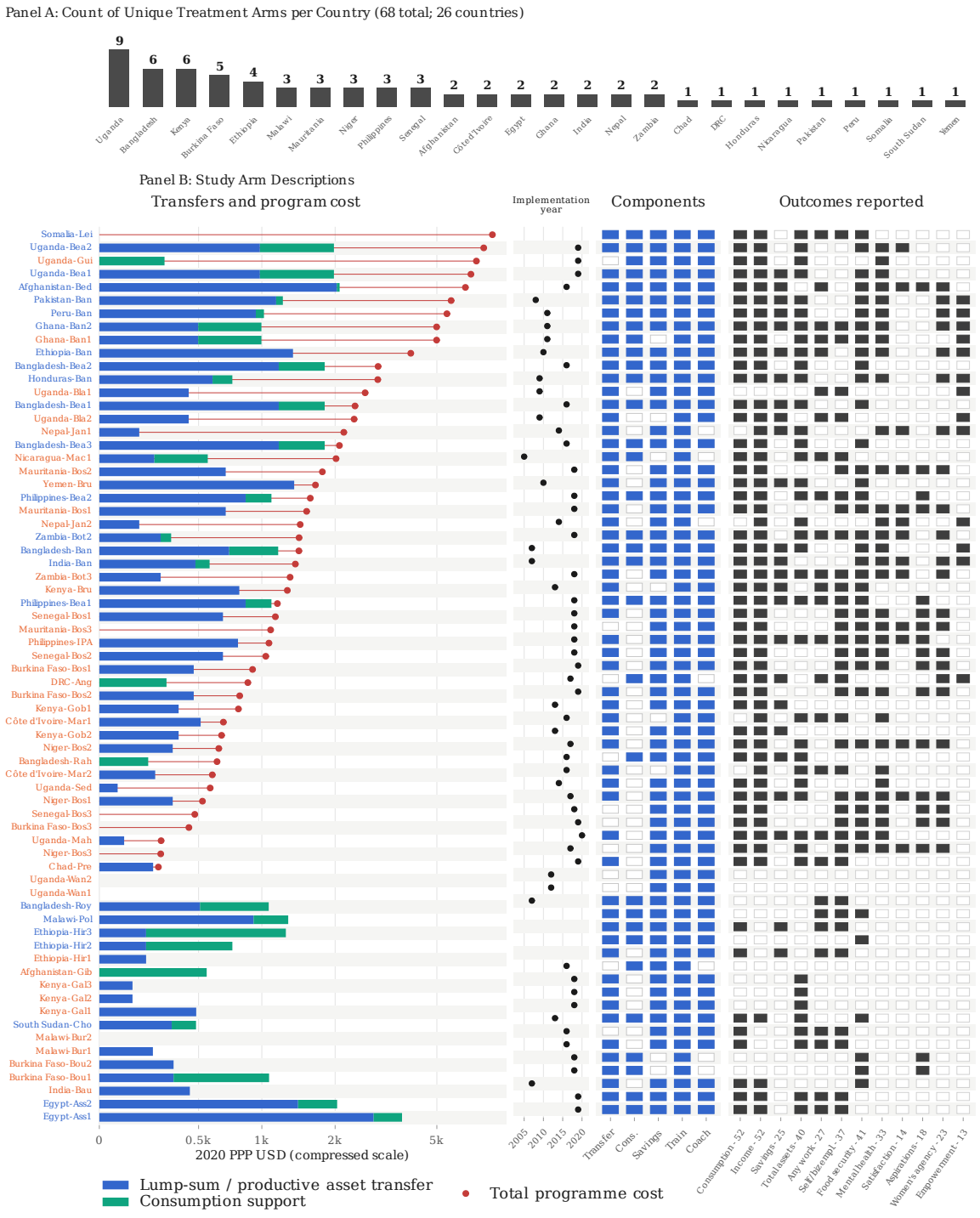
Romero, J., Esopo, K., McGuire, J., and Haushofer, J. The effect of economic transfers on psychological well-being and mental health. Working paper, 2021. URL https://johanneshaushofer.com/publications/Romero_et_al_Metaanalysis_2021.pdf.

Rubin, D. B. Estimation in Parallel Randomized Experiments. 6(4):377–401, 1981. ISSN 0362-9791. doi: 10.2307/1164617. URL https://www.jstor.org/stable/1164617.

Sedlmayr, R., Shah, A., and Sulaiman, M. Cash-plus: Poverty impacts of alternative transfer-based approaches. *Journal of Development Economics*, 144:102418, 2020. doi: 10.1016/j.jdeveco.2019.102418. Originally circulated as CSAE Working Paper WPS/2017-15 (2018) under the title "Cash-Plus: Poverty Impacts of Transfer-Based Intervention Alternatives".

Vivalt, E. How Much Can We Generalize From Impact Evaluations? 18(6):3045–3089, 2020. ISSN 1542-4766. doi: 10.1093/jeea/jvaa019. URL https://doi.org/10.1093/jeea/jvaa019.

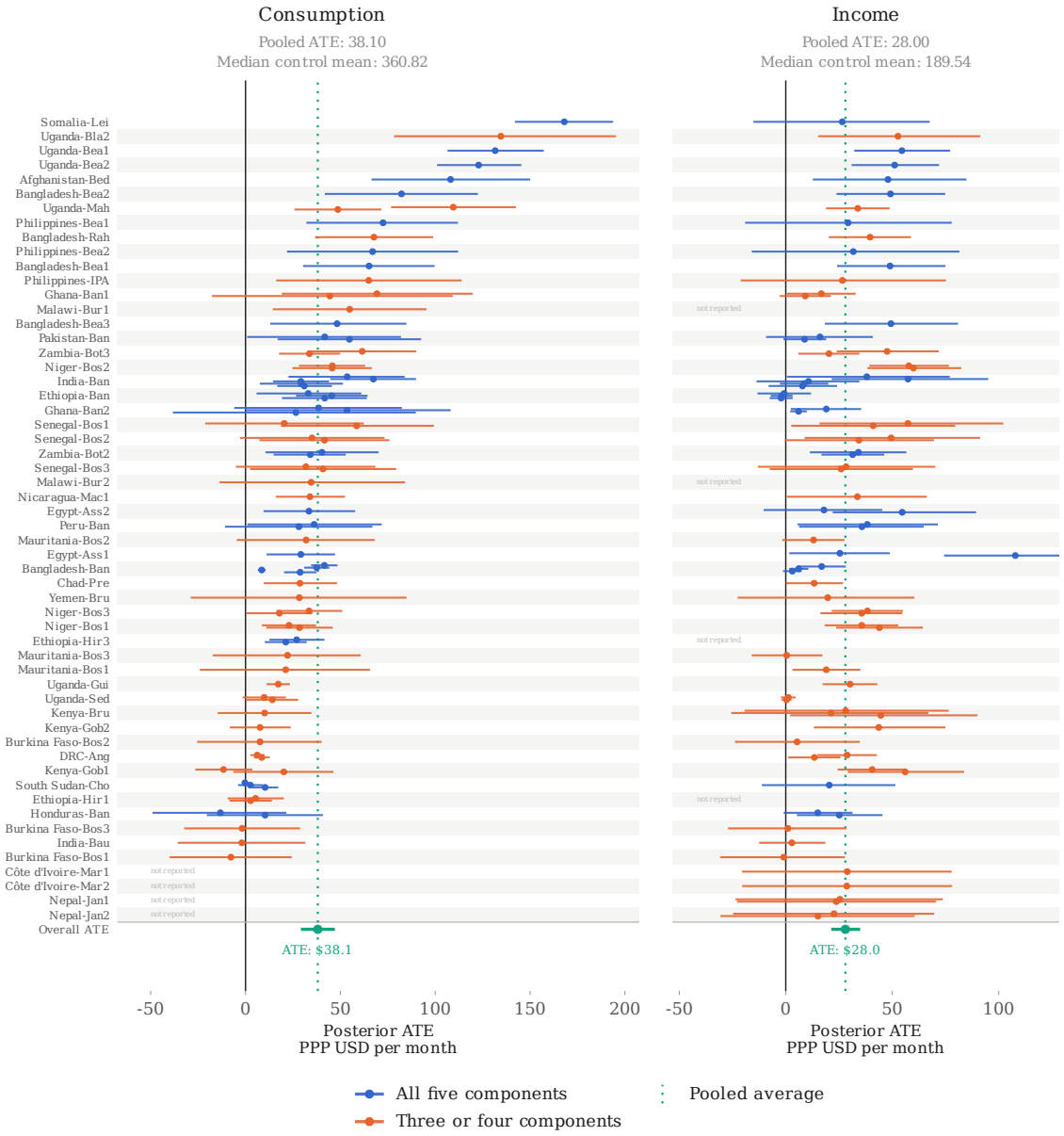
Figure 1: Programs, Components, Costs, and Outcome Coverage in the Evidence Base



Arm labels: blue = five components; orange = three or four; gray = count incomplete. ? = component unknown; missing costs are blank.

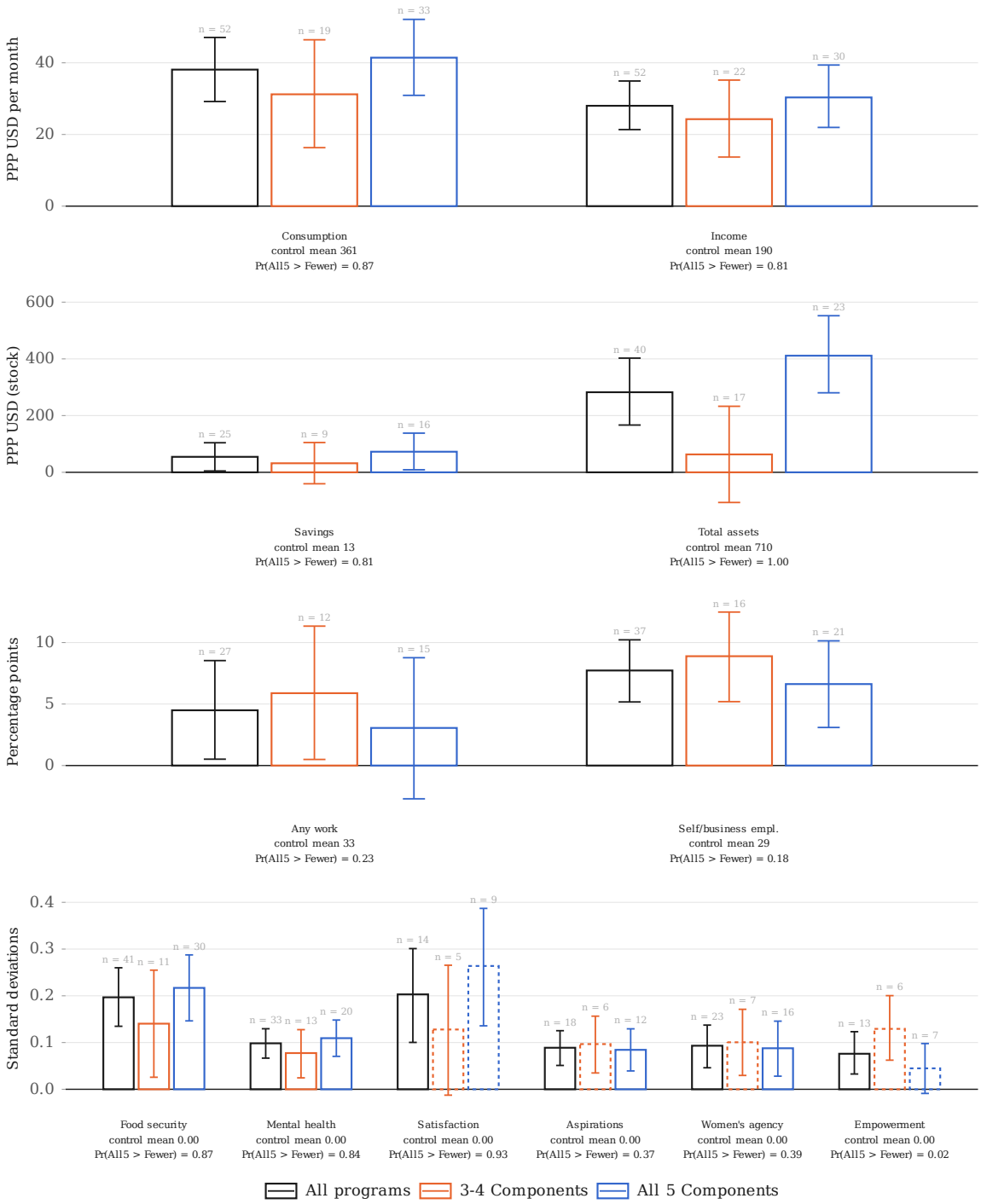
Notes: Each row represents one of 68 unique treatment arms. These consolidate 81 treatment variations reported across 49 study records from 41 publications, reports, and working papers. Across source reports, repeated follow-up measurements yield 123 treatment-variation-wave observations; after lineage pooling and harmonizing equivalent follow-up labels, the analysis contains 109 canonical treatment-arm-wave observations. Filled component and outcome cells indicate presence and reporting. Transfer amounts and total program costs are reported in 2020 PPP USD. Consumption-support amounts measure support additional to that received by the control group. For Ethiopia, Hirvonen T3 includes the additional maternal cash transfer, and T2+T3 denotes the pooled treatment contrast; T2 alone excludes support shared with the control.

Figure 2: Treatment-Arm Effects on Consumption and Income



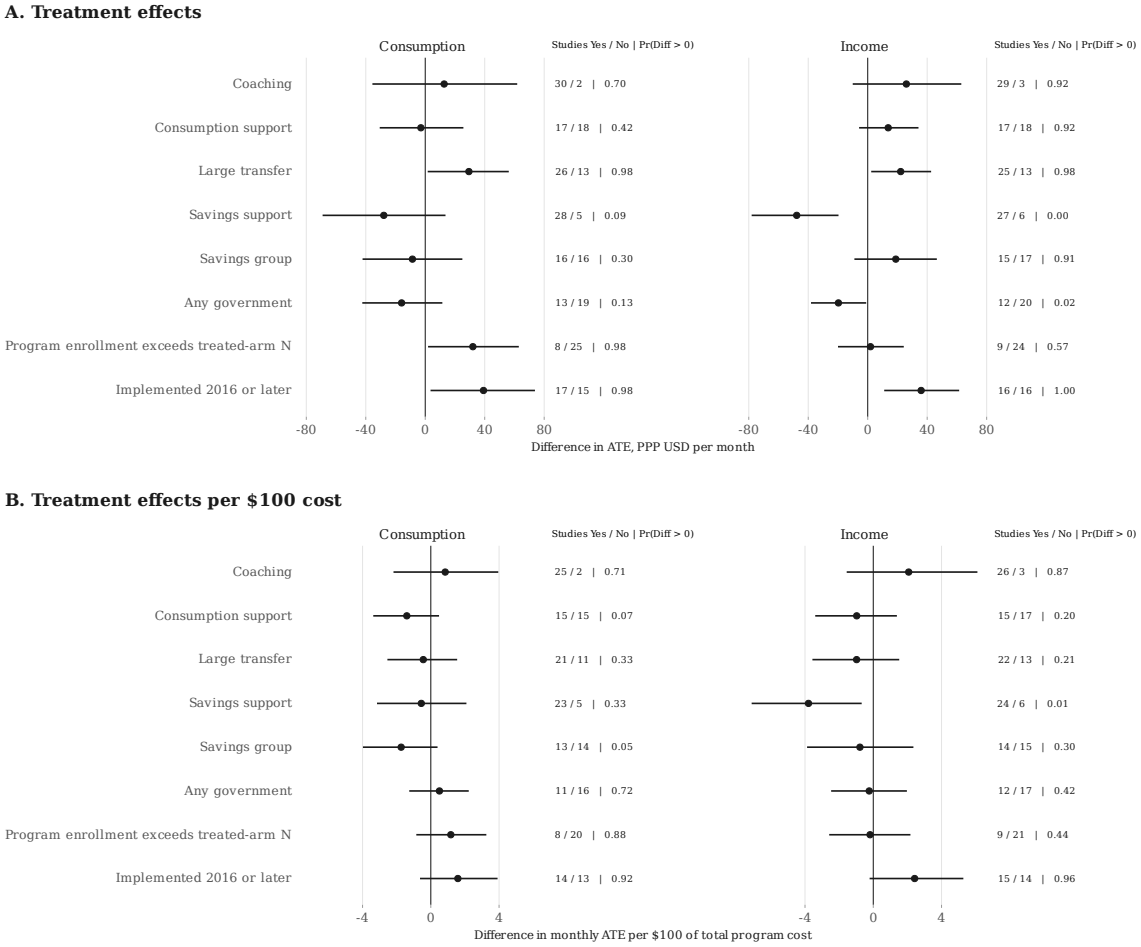
Rows are treatment arms and use a common ordering across panels. Points and horizontal lines report treatment-arm posterior estimates and 95 percent credible intervals. The solid black rule marks zero, and the dotted green rule marks the pooled average. The green Overall ATE row at the bottom of each panel reports the pooled average and its 95 percent credible interval.

Figure 3: Average Treatment Effects Across Outcomes and Program Composition



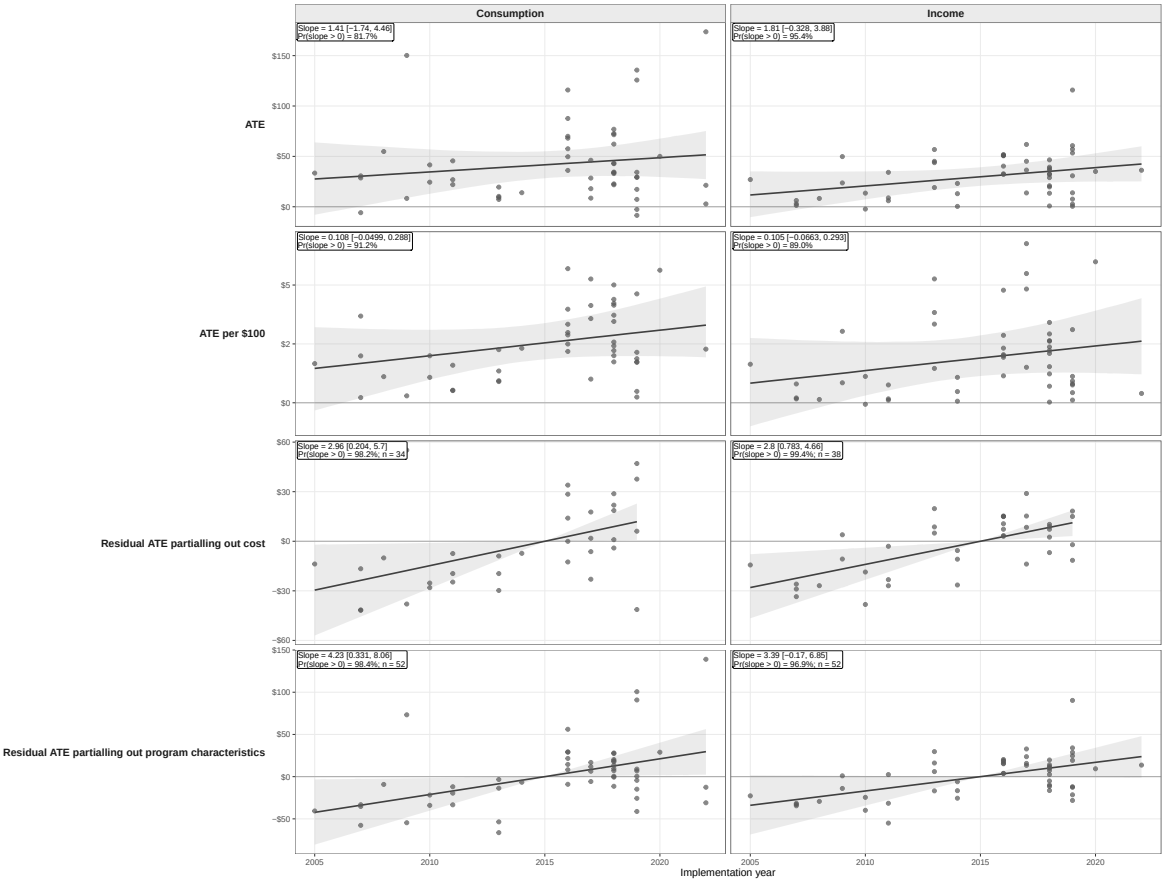
Notes: Bars report posterior medians and 95 percent credible intervals. Estimates are separated by measurement unit; dotted outlines mark estimates supported by seven or fewer studies. Appendix Table A.4, Panel A, compares the pooled consumption and income effects with a specification adding a shared randomization-level intercept.

Figure 4: Treatment Effects and Effects per $100 of Cost by Program Components and Characteristics



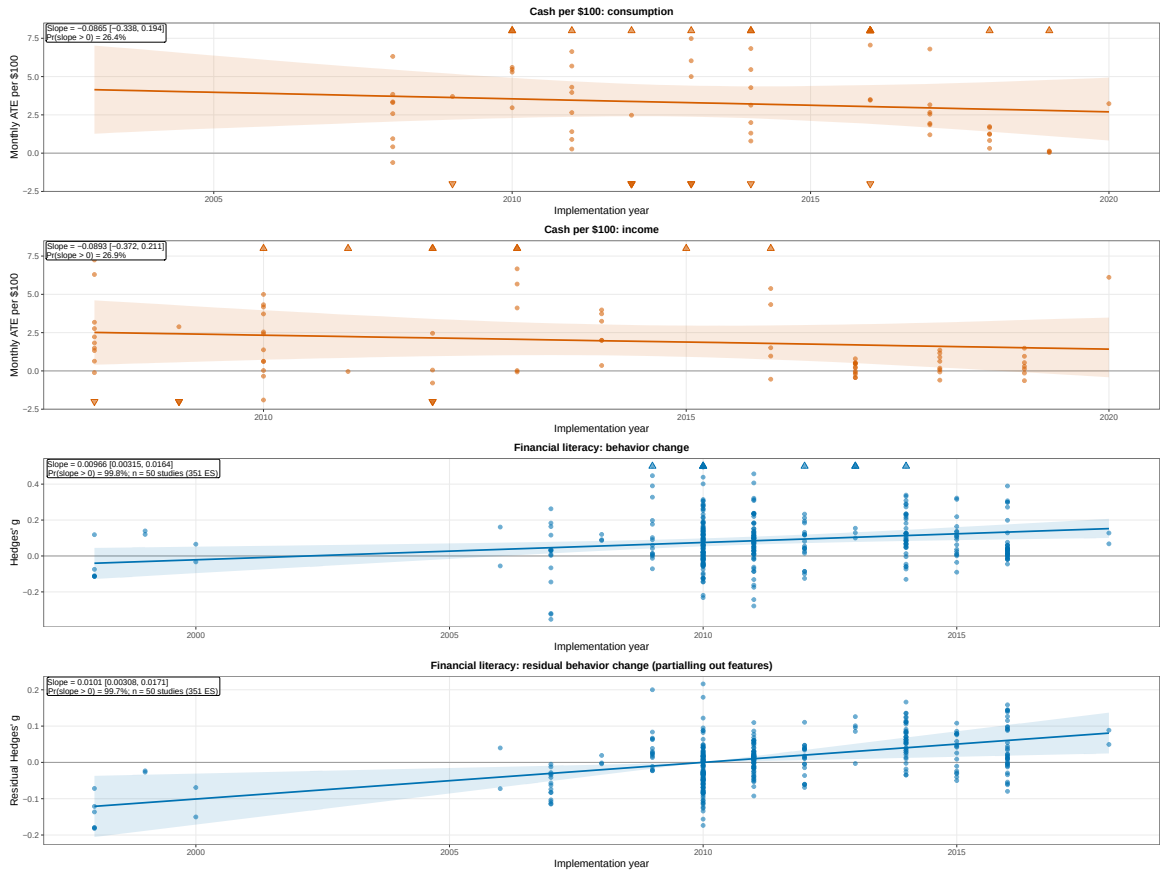
Notes: Panel A reports adjusted differences in monthly treatment effects; Panel B reports adjusted differences in monthly treatment effects per $100 of reported total program cost. Panel A includes 52 consumption and 52 income arms; Panel B includes the 44 consumption and 48 income arms with available positive cost estimates. Effects and costs are in 2020 PPP USD. Each arm's effect and standard error are divided by its cost before fitting the same simultaneous additive meta-regression. Missing costs are not imputed and cost uncertainty is not propagated. Counts refer to unique randomizations. Points report adjusted differences with minus without each program feature. Coaching, savings support, and consumption support are coded as components added relative to the named control; transfer size and the remaining characteristics are treatment-arm measures. "Large transfer" denotes a total productive plus consumption-support transfer of at least $400 per treated household in 2020 PPP USD. "Any government" includes any host-government involvement, including oversight or contracting. The implementation-era estimate compares programs implemented in 2016 or later with programs implemented through 2015. Whiskers are 95 percent credible intervals.

Figure 5: Treatment Effects and Program Implementation Year, Graduation Programs



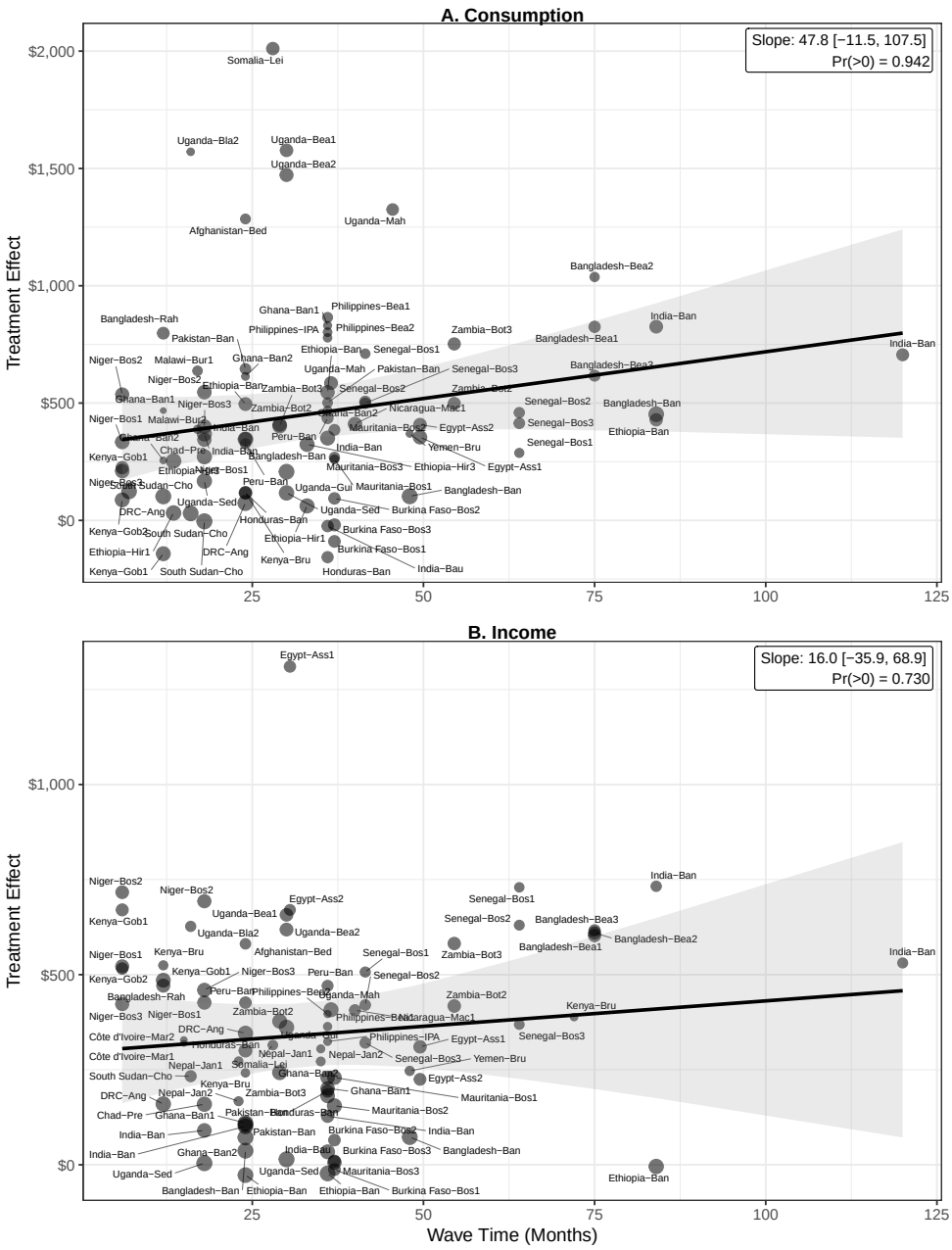
Notes: Columns show consumption and income. Rows report raw Graduation average treatment effects, treatment effects per $100 of total program cost, and treatment effects adjusted for predicted program cost and program characteristics, respectively. For the value-for-money row, the denominator is reported cost in 2020 PPP USD (including costs converted from nominal currency) where available, with component-only model predictions used otherwise; observations without a positive finite cost from either source are excluded. The cost-adjusted row instead controls for full-model predicted cost in 2020 PPP USD and is restricted to observations with that prediction and the required covariates. In the third and fourth rows, points show posterior median treatment effects adjusted to the sample means of the controls in each estimation sample. Within each posterior draw, the adjustment subtracts the fitted contribution of each control's deviation from its sample mean, retaining the estimated level. The program-characteristic controls are a large-transfer indicator, consumption support, coaching, savings support, and an organized savings-group indicator; training is excluded because it is universal in these estimation samples. Lines show posterior median implementation-year trends at these reference values, and shaded bands show pointwise 95 percent credible intervals for the mean, including uncertainty in its level. The bands are not prediction intervals for individual programs. Annotations report slopes, 95 percent credible intervals, and posterior probabilities of positive slopes. All specifications use the first follow-up for each treatment arm. Monetary effects are monthly household effects in 2020 PPP USD. Appendix Table A.4, Panel B, reports the corresponding randomization-level random-intercept sensitivity analysis for the raw effects in the top row.

Figure 6: Treatment Effects and Program Implementation Year, Cash and Financial Literacy



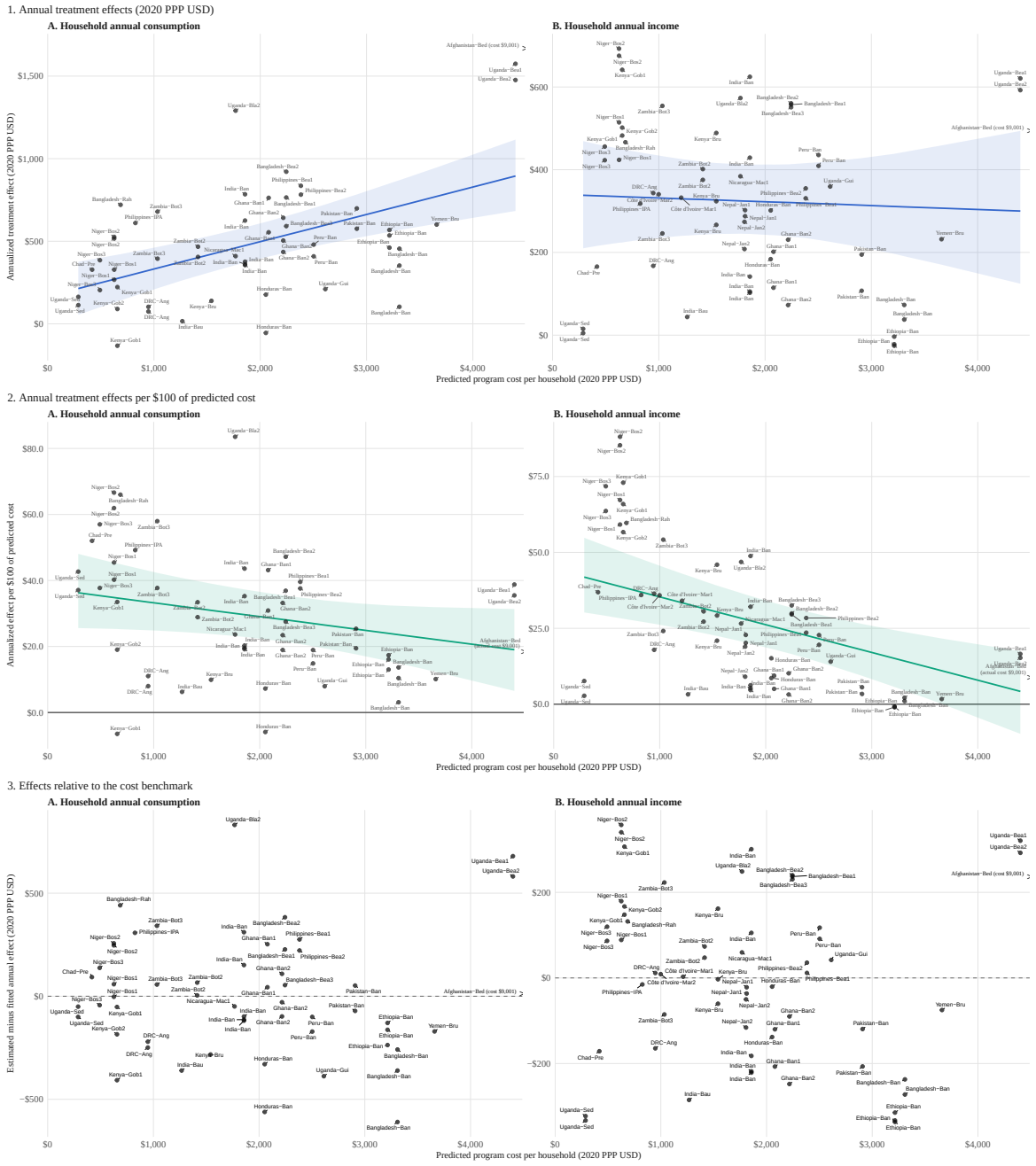
Notes: The first two rows report monthly cash transfer effects on consumption and income per $100 transferred. The third row reports standardized financial literacy effects on financial behavior in Hedges' g, controlling for follow-up timing. The fourth row reports residual financial literacy effects after additionally partialling out instruction intensity, participant age group, low-income targeting, developing-economy status, intervention channel, financial behavior outcome domain, and publication in a top economics journal. The financial literacy specifications use the earliest follow-up for 351 effect sizes from 50 studies in Kaiser et al. (2022b). Points outside the displayed cash-effect range or above 0.5 in the unadjusted financial literacy row are retained in the analysis and shown at the relevant boundary with triangles. Lines, credible bands, and annotations summarize hierarchical implementation-year meta-regressions; annotations report posterior median annual slopes, central 95 percent credible intervals in brackets, and posterior probabilities of a positive slope.

Figure 7: Persistence of Consumption and Income Treatment Effects



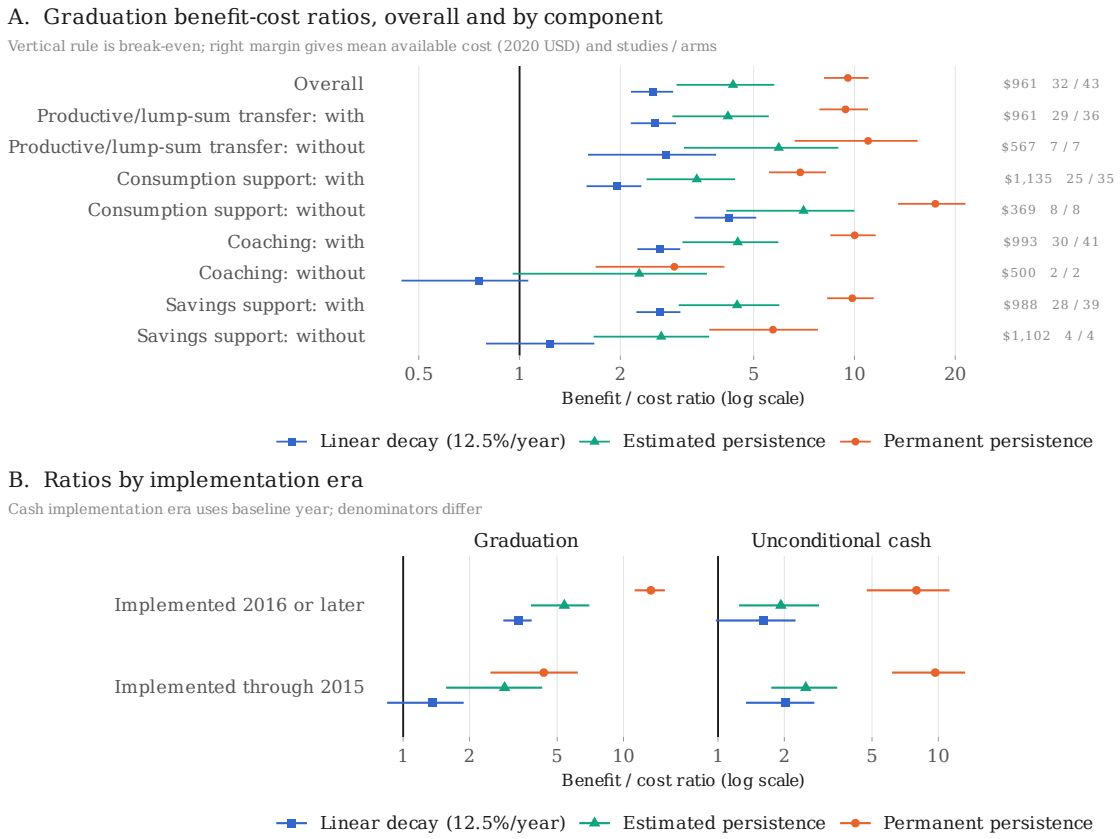
Notes: Panels A and B report observed treatment-arm follow-up effects for annual household consumption and income, respectively, in 2020 PPP USD. Larger points denote estimates with greater relative precision within each outcome panel. Lines and bands show fitted trends and 95 percent credible intervals. Boxes report annual slopes with 95 percent credible intervals in brackets and the probability of a positive slope.

Figure 8: Treatment Effects, Cost Effectiveness, and Value-Added Relative to Predicted Program Cost



Notes: Rows 1, 2, and 3 show annual treatment effect estimates, estimates per $100 of predicted cost, and estimated effects relative to the fitted cost benchmark, respectively; columns show consumption and income. The effect estimates incorporate hierarchical shrinkage. Row 2 retains its separately estimated cost-normalized specification. Row 3 shows each estimated effect in row 1 relative to the fitted effect at the same cost; uncertainty in these differences is not displayed. The dashed line marks effects equal to the cost benchmark. Lines and shaded regions in rows 1 and 2 retain the fitted relationships and 95 percent credible bands from those specifications. Predicted cost comes from the full cost model using available cost metadata and program components, for every included arm; reported costs are not substituted for predictions in this figure. Arms without the required predicted cost are excluded. Costs and annual effects are expressed in 2020 PPP USD. The horizontal axis is shortened to show differences among most programs more clearly; higher-cost programs remain in the analysis and appear at the right boundary with their actual costs.

Figure 9: Benefit-Cost Ratios for Graduation Programs



Notes: Panel A divides Graduation consumption benefits by reported full program cost in constant 2020 PPP USD. Arms without reported PPP costs are excluded from Graduation Benefit-Cost averages; predicted costs are not substituted. Right-margin costs are descriptive means in constant 2020 nominal USD across the same underlying programs represented in each return group, giving each program equal weight and using reported nominal costs where available and PPP-backconverted nominal equivalents otherwise; the overall mean is $961 and is available for all 31 programs. Panel B compares Graduation with unconditional cash within implementation era under matched linear decay, estimated persistence, and permanent persistence assumptions. Earlier programs began through 2015 and recent programs began in 2016 or later; this split is unchanged. Cash era uses baseline year because the UCT data do not report a separate intervention-start year. The cash denominator is the transfer plus an administrative-cost allowance equal to 24 percent of the transfer, following Table 11 of Crosta et al. (2026); the Graduation analytical denominator remains full program cost in constant 2020 PPP USD. Ratios use a log scale and give each underlying program equal weight; the vertical line marks break-even. Estimated persistence projects benefits over a common 10-year horizon, integrating over the full posterior distribution of the corresponding pooled consumption follow-up slope and flooring effects at zero. Permanent persistence holds the final observed annual effect constant forever. All scenarios discount benefits and costs at 5 percent per year. Component and implementation-era comparisons are descriptive rather than causal.

Appendix Tables

Table A.1: Support for the Core Program Components and Sample Sizes by Outcome

Panel A: Support for the Core Program Components

Component	All Graduation arms			Consumption		Income	
	With (1)	Without (2)	Within-experiment (3)	With (4)	Without (5)	With (6)	Without (7)
Productive/lump-sum transfer	57 / 42	11 / 10	6	44 / 35	8 / 8	45 / 35	7 / 7
Consumption support	49 / 34	19 / 14	1	43 / 30	9 / 9	39 / 28	13 / 11
Training	68 / 47	0 / 0	0	52 / 38	0 / 0	52 / 38	0 / 0
Coaching	61 / 42	7 / 5	0	50 / 36	2 / 2	48 / 35	4 / 3
Savings support	60 / 41	8 / 6	2	48 / 34	4 / 4	46 / 33	6 / 5
Unknown component status (arms / studies)							
Productive/lump-sum transfer		0 / 0			0 / 0		0 / 0
Consumption support		0 / 0			0 / 0		0 / 0
Training		0 / 0			0 / 0		0 / 0
Coaching		0 / 0			0 / 0		0 / 0
Savings support		0 / 0			0 / 0		0 / 0

Panel B: Sample Sizes by Outcome

Outcome	Studies (1)	Arms (2)	Arm-waves (3)
Consumption (HH Monthly, 2020 PPP USD)	38	52	83
Income (HH Monthly, 2020 PPP USD)	38	52	82
Food Security Index	29	41	61
Total Assets (Stock, 2020 PPP USD)	29	40	60
Mental Health Index	22	33	49
Self/Business Employment (Binary)	24	37	48
Any Work Reported (Binary)	22	27	33
Women's Agency Index	14	23	32
Savings (Stock, 2020 PPP USD)	24	25	30
Aspirations/Expectations Index	9	18	24
Empowerment Index	12	13	22
Satisfaction/Happiness Index	8	14	21

Notes: Panel A cells marked "With" and "Without" report unique treatment arms / canonical study records after consolidating repeated reports of the same intervention lineage across publications and follow-ups. "Within-experiment" is the number of randomizations containing included treatment arms on both sides of the component indicator, including arms recorded under different canonical study IDs within the same randomization. Productive/lump-sum transfer is a substantial productive asset or capital transfer. Consumption support is a transfer intended to meet short-run consumption needs. Training is organized instruction with identifiable content; coaching is recurring, application-oriented mentoring or advisory contact. Savings support includes program activities that facilitate or actively encourage saving, including organized groups, account access, savings targets, and structured encouragement. The consumption and income columns restrict to unique arms contributing those outcomes. Panel B reports canonical study records, unique treatment arms, and treatment-arm survey waves after the same lineage consolidation.

Appendix Tables

Table A.2: Cost Regressions (Unit and Total Cost, Constant 2020 Nominal USD)

Dependent Variables:	Log unit cost (constant 2020 nominal USD)	Log total cost (constant 2020 nominal USD)
Model:	(1)	(2)
Variables		
Constant	5.68***	10.6***
	(0.796)	(1.29)
Log(program size)	0.015	0.153
	(0.112)	(0.184)
N in program > N treated	0.407	-0.048
	(0.314)	(0.487)
Pre-existing infrastructure	-1.15***	-0.779
	(0.262)	(0.597)
Low data quality	-0.343	1.45*
	(0.234)	(0.675)
Moderate data quality	0.127	0.381
	(0.309)	(0.501)
Total transfer (00s)	0.057*	0.037
	(0.029)	(0.069)
Savings support	-0.115	0.307
	(0.270)	(0.584)
Coaching	0.163	0.942
	(0.239)	(0.591)
Consumption support	0.342	-0.250
	(0.311)	(0.675)
Fit statistics		
Observations	38	23
R^2	0.55605	0.29496
Adjusted R^2	0.41335	-0.19315

Heteroskedasticity-robust standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: Dependent variables are per-household unit cost and total program cost in constant 2020 nominal USD. Reported nominal costs are used where available; PPP-only source costs are back-converted to nominal equivalents using the study-country and cost-year exchange-rate relationship. Total transfer enters as hundreds of 2020 PPP USD because it measures comparable treatment intensity across settings. Regressions use observations with available dependent variables and required covariates; predicted costs do not replace missing dependent-variable costs. Standard errors are heteroskedasticity robust.

Appendix Tables

Table A.3: Overall Treatment Effects

Outcome	Median Control Mean (1)	Treatment Effect [95% Credibility Interval] Overall ATE (2)	3–4 Components (3)	All 5 Components (4)	Difference: (4) − (3) (5)	Studies / arms (6)
Consumption (HH Monthly)	361	38*** [29, 47]	31*** [16, 46]	41*** [31, 52]	10 [-8, 28] $Pr(\Delta > 0)$=0.87	Col 2: 38 / 52 Col 3: 19 / 19 Col 4: 25 / 33
Income (HH Monthly)	190	28*** [21, 35]	24*** [14, 35]	30*** [22, 39]	6 [-8, 20] $Pr(\Delta > 0)$=0.81	Col 2: 38 / 52 Col 3: 20 / 22 Col 4: 23 / 30
Food Security (Index)	0.00	0.20*** [0.13, 0.26]	0.14** [0.03, 0.25]	0.22*** [0.15, 0.29]	0.08 [-0.06, 0.21] $Pr(\Delta > 0)$=0.87	Col 2: 29 / 41 Col 3: 10 / 11 Col 4: 24 / 30
Savings (Stock)	13	55** [5, 104]	32 [-40, 105]	73** [9, 138]	41 [-53, 135] $Pr(\Delta > 0)$=0.81	Col 2: 24 / 25 Col 3: 9 / 9 Col 4: 15 / 16
Total Assets (Stock)	710	282*** [167, 403]	63 [-106, 233]	411*** [280, 552]	349*** [136, 564] $Pr(\Delta > 0)$=1.00	Col 2: 29 / 40 Col 3: 13 / 17 Col 4: 19 / 23
Any Work Reported (Binary)	0.33	0.04** [0.01, 0.09]	0.06** [0.00, 0.11]	0.03 [-0.03, 0.09]	-0.03 [-0.11, 0.05] $Pr(\Delta > 0)$=0.23	Col 2: 22 / 27 Col 3: 11 / 12 Col 4: 13 / 15
Self/Business Employment (Binary)	0.29	0.08*** [0.05, 0.10]	0.09*** [0.05, 0.12]	0.07*** [0.03, 0.10]	-0.02 [-0.07, 0.03] $Pr(\Delta > 0)$=0.18	Col 2: 24 / 37 Col 3: 15 / 16 Col 4: 15 / 21
Mental Health (Index)	0.00	0.10*** [0.07, 0.13]	0.08*** [0.02, 0.13]	0.11*** [0.07, 0.15]	0.03 [-0.03, 0.10] $Pr(\Delta > 0)$=0.84	Col 2: 22 / 33 Col 3: 11 / 13 Col 4: 16 / 20
Satisfaction/Happiness (Index)	0.00	0.20*** [0.10, 0.30]	0.13** [-0.01, 0.27]	0.26*** [0.14, 0.39]	0.14* [-0.05, 0.31] $Pr(\Delta > 0)$=0.93	Col 2: 8 / 14 Col 3: 4 / 5 Col 4: 7 / 9
Aspirations/Expectations (Index)	0.00	0.09*** [0.05, 0.13]	0.10*** [0.03, 0.16]	0.08*** [0.04, 0.13]	-0.01 [-0.08, 0.06] $Pr(\Delta > 0)$=0.37	Col 2: 9 / 18 Col 3: 5 / 6 Col 4: 8 / 12
Women's Agency (Index)	0.00	0.09*** [0.05, 0.14]	0.10*** [0.03, 0.17]	0.09*** [0.03, 0.15]	-0.01 [-0.10, 0.08] $Pr(\Delta > 0)$=0.39	Col 2: 14 / 23 Col 3: 7 / 7 Col 4: 12 / 16
Empowerment (Index)	0.00	0.08*** [0.03, 0.12]	0.13*** [0.06, 0.20]	0.04* [-0.01, 0.10]	-0.08** [-0.17, -0.00] $Pr(\Delta > 0)$=0.02	Col 2: 12 / 13 Col 3: 5 / 6 Col 4: 7 / 7

Notes: Column (1) reports the median control-group mean at baseline across contributing estimates. Column (2) reports the overall average treatment effect from the model without covariates. Columns (3) and (4) report standardized posterior predictions from a model including a binary covariate set to 1 if the arm has all 5 graduation components used for inclusion criteria, setting the all-five-components indicator to zero or one while averaging over the observed sample. Column (5) reports the difference between these predictions and the posterior probability that it is positive. Stars in columns (2)–(5) summarize the posterior probability that the corresponding estimate is positive: * denotes $Pr(\theta > 0) \geq 0.90$ or ≤ 0.10, ** denotes ≥ 0.95 or ≤ 0.05, and *** denotes ≥ 0.99 or ≤ 0.01. Column (6) reports canonical studies / unique treatment arms after consolidating repeated reports of the same intervention lineage across publications and follow-ups. All monetary figures reported in 2020 PPP USD. The qualifier beneath each outcome label gives the unit of measurement. Cells report posterior medians with 95 percent credible intervals in brackets.

Appendix Tables

Table A.4: Randomization-Level Random-Effects Robustness: Consumption and Income

Outcome	Headline specification	Randomization intercept
Panel A: Pooled monthly household effects		
Consumption	38.22 [29.32, 47.40]	42.97 [26.44, 59.95]
Income	28.08 [21.36, 34.88]	25.41 [16.59, 34.82]
Panel B: Annual implementation-year slopes		
Consumption	1.50 [-1.71, 4.51]	1.70 [-1.83, 5.32]
Income	1.80 [-0.41, 3.89]	1.82 [-0.61, 4.27]

Notes: Entries report posterior medians with central 95 percent credible intervals. The robustness specification adds a shared randomization-level intercept to the existing arm-wave residual heterogeneity; source publications from the same randomization share that intercept. Both columns use identical outcome-specific samples. Panel A retains all eligible follow-ups and reports monthly household effects in 2020 PPP USD. Panel B uses the earliest eligible follow-up for each pooled treatment arm and outcome, and reports the change in monthly household effects per one-year increase in implementation year. The randomization intercept has a zero-centered normal distribution; its standard deviation has the same outcome-scaled half-normal prior as residual heterogeneity. Fixed-effect priors and sample selection are unchanged. This is a sensitivity analysis for dependence across estimates, not a correction using known sampling-error covariances. The implementation-year comparison covers the raw effects in the top row of the main implementation-year figure; it does not refit the per-cost or component-adjusted rows.

Appendix Tables

Table A.5: Savings Robustness Excluding the Ethiopia Banerjee/Barker Estimate

Specification	Pooled ATE [95% credible interval] (1)
Baseline hierarchical model	54.52** [4.71, 104.22]
Excluding Ethiopia Banerjee/Barker	30.15*** [14.20, 46.05]

Notes: Both rows use the no-covariate hierarchical model. The second row excludes the Ethiopia Banerjee/Barker savings estimate. Heterogeneity SD measures dispersion in underlying effects; the mean pooling factor measures shrinkage toward the pooled mean, with higher values indicating greater pooling. Brackets report 95 percent credible intervals.

Appendix Tables

Table A.6: Heterogeneity, Pooling, and Component Decomposition

Outcome	No covariates			Coaching, savings, consumption support, and transfer size		
	Arm-wave SD ($\hat{\sigma}_\tau$) (1)	Mean pooling factor ($\bar{\omega}$) (2)	Global pooling factor (λ) (3)	Residual SD ($\hat{\sigma}_u$) (4)	Predicted-mean SD ($\widehat{\mathrm{SD}}(X\hat{\beta})$) (5)	Heterogeneity explained (R^2_{het}) (6)
Consumption (HH Monthly, 2020 PPP USD) 51 / 51 arms	37.10 [30.44, 44.96]	0.18 [0.14, 0.23]	0.19	36.39 [29.92, 44.52]	13.50 [6.20, 21.97]	0.12 [0.03, 0.27]
Income (HH Monthly, 2020 PPP USD) 50 / 50 arms	24.78 [19.46, 31.68]	0.35 [0.28, 0.41]	0.36	26.15 [20.45, 33.33]	8.39 [3.04, 16.39]	0.09 [0.01, 0.27]
Food Security Index 41 / 41 arms	0.233 [0.194, 0.286]	0.09 [0.06, 0.12]	0.09	0.233 [0.192, 0.287]	0.067 [0.026, 0.116]	0.08 [0.01, 0.20]
Savings (Stock, 2020 PPP USD) 24 / 24 arms	140.82 [108.38, 188.36]	0.01 [0.01, 0.02]	0.01	143.67 [110.16, 193.63]	45.78 [17.02, 90.36]	0.09 [0.01, 0.28]
Total Assets (Stock, 2020 PPP USD) 38 / 38 arms	445.56 [346.30, 573.95]	0.07 [0.05, 0.09]	0.07	407.40 [310.93, 534.14]	206.69 [109.27, 313.17]	0.21 [0.06, 0.39]
Any Work Reported (Binary) 26 / 26 arms	0.116 [0.091, 0.156]	0.04 [0.02, 0.07]	0.04	0.104 [0.080, 0.141]	0.0594 [0.0287, 0.0933]	0.24 [0.06, 0.47]
Self/Business Employment (Binary) 36 / 36 arms	0.086 [0.069, 0.110]	0.10 [0.07, 0.14]	0.11	0.0760 [0.0595, 0.0985]	0.0449 [0.0237, 0.0671]	0.26 [0.08, 0.47]
Aspirations/Expectations Index 18 / 18 arms	0.071 [0.046, 0.108]	0.41 [0.25, 0.60]	0.44	0.073 [0.047, 0.110]	0.0184 [0.0036, 0.0448]	0.06 [0.00, 0.29]
Mental Health Index 32 / 32 arms	0.095 [0.072, 0.126]	0.28 [0.20, 0.39]	0.29	0.093 [0.070, 0.125]	0.0413 [0.0174, 0.0701]	0.16 [0.03, 0.39]
Satisfaction/Happiness Index 14 / 14 arms	0.216 [0.154, 0.313]	0.13 [0.07, 0.22]	0.13	0.201 [0.143, 0.294]	0.113 [0.042, 0.196]	0.24 [0.04, 0.52]
Empowerment Index 13 / 13 arms	0.076 [0.039, 0.131]	0.43 [0.25, 0.71]	0.48	0.0243 [0.0011, 0.0723]	0.086 [0.054, 0.120]	0.93 [0.53, 1.00]
Women's Agency Index 23 / 23 arms	0.112 [0.081, 0.160]	0.27 [0.16, 0.39]	0.28	0.112 [0.080, 0.158]	0.0410 [0.0132, 0.0775]	0.12 [0.01, 0.35]

Notes: The component specification includes coaching, savings support, consumption support, and transfer size. $\hat{\sigma}_\tau$ is the between-program standard deviation in the no-covariate model. $\bar{\omega}$ is the mean arm-level pooling factor, with its 95 percent credible interval; λ is the Gelman–Pardoe global pooling factor. $\hat{\sigma}_u$ is the residual between-program standard deviation after accounting for the listed components, and $\widehat{\mathrm{SD}}(X\hat{\beta})$ is the standard deviation of component-predicted means across programs. $R^2_{het} = \widehat{\mathrm{Var}}(X\hat{\beta})/(\widehat{\mathrm{Var}}(X\hat{\beta}) + \hat{\sigma}^2_u)$ is the share of true between-program heterogeneity explained by the specification. The same outcome-specific arms contribute to both models. Substantial residual heterogeneity remains for most outcomes.

Appendix Tables

Table A.7: Adjusted effects by transfer size

Outcome	No/small < $400 (1)	Medium $400–799 (2)	Large ≥ $800 (3)	Pairwise $Pr(\Delta > 0)$
Consumption HH mo., PPP USD	**24.37** [9.82, 39.36]	**35.19** [21.24, 49.38]	**53.23** [38.52, 68.29]	Δ_{M-S}: 0.86 Δ_{L-S}: 1.00 Δ_{L-M}: 0.96
Income HH mo., PPP USD	**24.09** [11.82, 36.30]	**31.90** [19.53, 44.54]	**27.87** [16.79, 39.99]	Δ_{M-S}: 0.81 Δ_{L-S}: 0.67 Δ_{L-M}: 0.32
Food Security index (SD)	**0.20** [0.07, 0.33]	**0.18** [0.08, 0.27]	**0.22** [0.12, 0.31]	Δ_{M-S}: 0.36 Δ_{L-S}: 0.57 Δ_{L-M}: 0.73
Savings Stock, PPP USD	**22.70** [-52.62, 99.74]	**36.91** [-65.54, 137.82]	**98.06** [15.57, 181.21]	Δ_{M-S}: 0.59 Δ_{L-S}: 0.92 Δ_{L-M}: 0.83
Total Assets Stock, PPP USD	**42.80** [-155.24, 246.25]	**211.79** [3.40, 420.80]	**469.35** [309.59, 643.37]	Δ_{M-S}: 0.88 Δ_{L-S}: 1.00 Δ_{L-M}: 0.97
Any Work Reported binary	**-0.01** [-0.07, 0.06]	**0.08** [0.01, 0.14]	**0.06** [-0.01, 0.13]	Δ_{M-S}: 0.97 Δ_{L-S}: 0.93 Δ_{L-M}: 0.34
Self/Business Empl. binary	**0.05** [0.01, 0.10]	**0.09** [0.05, 0.13]	**0.09** [0.03, 0.15]	Δ_{M-S}: 0.92 Δ_{L-S}: 0.83 Δ_{L-M}: 0.45
Mental Health index (SD)	**0.08** [0.02, 0.13]	**0.09** [0.03, 0.14]	**0.13** [0.08, 0.18]	Δ_{M-S}: 0.56 Δ_{L-S}: 0.91 Δ_{L-M}: 0.88
Empowerment index (SD)	**0.21** [0.14, 0.27]	**0.01** [-0.08, 0.10]	**0.04** [0.00, 0.07]	Δ_{M-S}: 0.00 Δ_{L-S}: 0.00 Δ_{L-M}: 0.72
Women's Agency index (SD)	**0.10** [0.03, 0.16]	**0.10** [0.02, 0.17]	**0.07** [-0.04, 0.17]	Δ_{M-S}: 0.50 Δ_{L-S}: 0.31 Δ_{L-M}: 0.33

Notes: Bold values are posterior medians of treatment effects from a hierarchical meta-regression with an intercept and indicators for medium and large transfers; brackets contain 95% credible intervals. This specification does not additionally control for other program components. Transfer categories use total productive plus consumption-support transfers per treated household in 2020 PPP USD: no/small is below $400, medium is at least $400 but below $800, and large is at least $800. Outcome units appear below each outcome label; monetary outcomes are in 2020 PPP USD. In the final column, S denotes the pooled no/small category, M medium, and L large; the entries are posterior probabilities that the indicated difference is positive (M−S, L−S, and L−M). Dashes denote comparisons not identified by the available transfer categories in the outcome-specific estimation sample.

Appendix Tables

Table A.8: Transfer-Size Robustness Including Partly Repayable Grants

Panel A: Combined medium/large-transfer specification		
Outcome	Excluding repayable arms (1)	Including repayable arms (2)
Consumption	**18.09**	**18.58**
HH mo., PPP USD	[-4.23, 40.54]	[-3.33, 40.56]
Income	**4.02**	**4.58**
HH mo., PPP USD	[-14.81, 24.01]	[-14.55, 24.15]
Food Security	**0.02**	**0.02**
index (SD)	[-0.14, 0.18]	[-0.14, 0.18]
Savings	**41.35**	**43.54**
stock, PPP USD	[-103.53, 187.18]	[-101.26, 187.85]
Total Assets	**235.18**	**226.12**
stock, PPP USD	[-88.68, 563.27]	[-83.81, 539.00]
Any Work Reported	**0.12**	**0.11**
binary	[0.02, 0.21]	[0.03, 0.20]
Self/Business Empl.	**0.06**	**0.06**
binary	[0.00, 0.11]	[0.01, 0.11]
Aspirations/Expectations	**0.00**	**0.00**
index (SD)	[-0.07, 0.07]	[-0.07, 0.07]
Mental Health	**0.01**	**0.01**
index (SD)	[-0.07, 0.08]	[-0.06, 0.08]
Satisfaction/Happiness	**0.03**	**0.03**
index (SD)	[-0.18, 0.24]	[-0.18, 0.24]
Women's Agency	**0.02**	**0.02**
index (SD)	[-0.08, 0.12]	[-0.08, 0.12]

Panel B: Separate medium- and large-transfer specification				
	Medium transfer		Large transfer	
Outcome	Excluding (1)	Including (2)	Excluding (3)	Including (4)
Consumption	**10.78**	**12.15**	**28.82**	**29.03**
HH mo., PPP USD	[-9.60, 30.80]	[-8.01, 32.23]	[8.11, 49.60]	[7.92, 49.88]
Income	**7.76**	**8.04**	**3.69**	**3.53**
HH mo., PPP USD	[-9.09, 25.34]	[-8.63, 25.15]	[-12.16, 21.08]	[-12.66, 20.69]
Food Security	**-0.03**	**-0.03**	**0.01**	**0.01**
index (SD)	[-0.18, 0.13]	[-0.18, 0.13]	[-0.14, 0.17]	[-0.14, 0.17]
Savings	**14.06**	**17.97**	**75.02**	**74.41**
stock, PPP USD	[-109.43, 132.46]	[-97.20, 132.88]	[-34.24, 183.80]	[-32.15, 176.96]
Total Assets	**169.74**	**190.67**	**426.85**	**426.62**
stock, PPP USD	[-118.25, 448.72]	[-80.96, 457.74]	[174.20, 683.70]	[183.58, 677.53]
Any Work Reported	**0.09**	**0.09**	**0.07**	**0.07**
binary	[-0.00, 0.18]	[-0.00, 0.17]	[-0.02, 0.16]	[-0.02, 0.15]
Self/Business Empl.	**0.04**	**0.04**	**0.03**	**0.04**
binary	[-0.02, 0.09]	[-0.01, 0.09]	[-0.04, 0.10]	[-0.03, 0.10]
Aspirations/Expectations	**0.00**	**0.00**	**-0.00**	**-0.00**
index (SD)	[-0.07, 0.07]	[-0.07, 0.07]	[-0.11, 0.11]	[-0.11, 0.11]
Mental Health	**0.01**	**0.01**	**0.05**	**0.05**
index (SD)	[-0.07, 0.08]	[-0.07, 0.08]	[-0.02, 0.13]	[-0.02, 0.12]
Satisfaction/Happiness	**0.05**	**0.05**	**0.21**	**0.21**
index (SD)	[-0.13, 0.24]	[-0.13, 0.24]	[-0.04, 0.44]	[-0.04, 0.44]
Empowerment	**-0.20**	**-0.20**	**-0.17**	**-0.17**
index (SD)	[-0.30, -0.09]	[-0.30, -0.09]	[-0.24, -0.09]	[-0.24, -0.09]
Women's Agency	**-0.00**	**-0.00**	**-0.03**	**-0.03**
index (SD)	[-0.09, 0.09]	[-0.09, 0.09]	[-0.15, 0.09]	[-0.15, 0.09]

Notes: The baseline transfer-moderator specifications exclude two arms whose productive transfer or cash grant is partly repayable: the hybrid Ultra Poor Graduation arm in Rahman et al. (study ID RAH2021_8765) and the "Cash grant + repayment" arm in MAR2024_5618. Panel A combines medium and large transfers in one indicator; Panel B estimates separate indicators for medium and large transfers. Within each panel, the inclusion specification adds both partly repayable arms back while leaving the model, outcomes, and all other sample restrictions unchanged. Point estimates are bold; 95 percent credible intervals appear in gray brackets.

Appendix Tables

Table A.9: Treatment Effects with and without Program Components and Characteristics

Program feature	Consumption			Income		
	Without (1)	With (2)	Difference (3)	Without (4)	With (5)	Difference (6)
Panel A: Graduation inclusion components						
Coaching	31.32 [-16.32, 78.82]	44.09 [32.63, 55.78]	12.70 [-35.53, 61.86] $Pr(\Delta > 0)$=0.70	5.53 [-30.39, 41.10]	31.50 [23.63, 40.36]	25.99* [-10.07, 62.95] $Pr(\Delta > 0)$=0.92
Savings support	68.25 [29.53, 106.98]	40.42 [28.20, 52.54]	-27.89* [-69.04, 13.63] $Pr(\Delta > 0)$=0.09	69.96 [43.75, 98.45]	22.14 [13.94, 31.02]	-47.76*** [-78.06, -19.64] $Pr(\Delta > 0)$=0.00
Consumption support	44.84 [28.67, 61.18]	41.97 [22.02, 62.12]	-2.93 [-30.61, 25.66] $Pr(\Delta > 0)$=0.42	23.71 [11.33, 36.43]	37.34 [24.96, 51.16]	13.78* [-5.79, 34.15] $Pr(\Delta > 0)$=0.92
Large transfer	22.86 [1.03, 44.89]	52.10 [37.82, 66.07]	29.36** [1.65, 56.19] $Pr(\Delta > 0)$=0.98	14.12 [-1.81, 30.60]	36.32 [26.29, 47.29]	22.18** [2.36, 42.65] $Pr(\Delta > 0)$=0.98
Panel B: Other program characteristics						
Savings group	48.74 [26.01, 72.27]	40.20 [22.72, 57.35]	-8.62 [-42.16, 24.97] $Pr(\Delta > 0)$=0.30	18.96 [0.66, 38.23]	37.92 [24.59, 51.65]	18.89* [-8.95, 46.50] $Pr(\Delta > 0)$=0.91
Any government	51.50 [34.53, 68.68]	35.59 [17.96, 54.17]	-15.90 [-42.33, 11.51] $Pr(\Delta > 0)$=0.13	38.60 [27.81, 50.31]	18.94 [5.62, 32.77]	-19.70** [-38.13, -1.02] $Pr(\Delta > 0)$=0.02
Program enrollment exceeds treated-arm N	36.25 [22.76, 49.48]	68.27 [42.30, 94.47]	32.03** [1.91, 62.93] $Pr(\Delta > 0)$=0.98	28.94 [19.70, 39.09]	30.89 [12.35, 50.20]	1.87 [-19.98, 24.25] $Pr(\Delta > 0)$=0.57
Implemented in 2016 or later	17.30 [-8.70, 43.96]	56.41 [40.36, 72.91]	39.18** [3.56, 73.79] $Pr(\Delta > 0)$=0.98	6.76 [-10.29, 23.80]	42.67 [30.21, 55.98]	35.94*** [11.10, 61.51] $Pr(\Delta > 0)$=1.00

Notes: Posterior predictions from the simultaneous additive meta-regression in Figure 4. Without and With set the indicated feature to zero and one, averaging the other covariates over the fitted sample; differences are With minus Without. Coaching, savings support, and consumption support measure components added relative to the named control. Savings support includes groups, account access, savings targets, and structured encouragement; organized savings groups also enter separately. Large transfer denotes total productive plus consumption-support transfers of at least $400 per treated household in 2020 PPP USD. Any government includes host-government oversight or contracting. Program enrollment is compared with the evaluated treatment-arm sample; unresolved comparisons are coded zero. The implementation indicator compares programs beginning in 2016 or later with earlier programs. Each arm contributes its first available follow-up. Cells report posterior medians and 95 percent credible intervals in brackets. Stars apply to differences and follow the posterior-probability thresholds in Appendix Table A.3.

Appendix Tables

Table A.10: Value-for-Money Trends by Main Outcome

Outcome	Posterior median (1)	95% credible interval (2)	Pr(slope > 0) (3)
Consumption			
Consumption (HH Monthly, 2020 PPP USD)	0.108	[-0.0499, 0.288]	91.2%
Income			
Income (HH Monthly, 2020 PPP USD)	0.105	[-0.0663, 0.293]	89.0%
Food Security			
Food Security Index	0.000766	[-0.000588, 0.00219]	87.3%
Assets			
Savings (Stock, 2020 PPP USD)	0.0344	[-0.539, 0.651]	55.0%
Total Assets (Stock, 2020 PPP USD)	0.556	[-1.02, 2.2]	76.2%
Employment			
Any Work Reported (Binary)	-0.000679	[-0.00122, -0.000135]	0.9%
Self/Business Employment (Binary)	-0.000413	[-0.00104, 0.000218]	9.2%
Psychological			
Mental Health Index	0.000209	[-0.000524, 0.00102]	72.1%
Satisfaction/Happiness Index	0.0011	[-0.0046, 0.00738]	65.5%
Aspirations/Expectations Index	0.000363	[-0.00778, 0.00918]	53.6%
Empowerment			
Women's Agency Index	0.000383	[-0.00185, 0.00253]	64.1%
Empowerment Index	0.00166	[0.000162, 0.00332]	98.2%

Notes: Outcomes are restricted to the main outcomes defined for the paper. Annual slopes come from unconditional implementation-year value-for-money meta-regressions using reported 2020 PPP USD program costs (including costs converted from nominal currency) where available and component-only model predictions otherwise. Observations without a positive finite cost from either source are excluded. The table reports posterior median annual slopes, central 95 percent credible intervals, and posterior probabilities of a positive slope.

Appendix Tables

Table A.11: Average Benefit/Cost Ratios Overall and by Program Component

Group	Linear decay		Estimated persistence		Permanent persistence			
	B/C (1)	Pr($B/C > 1$) (2)	B/C (3)	Pr($B/C > 1$) (4)	B/C (5)	Pr($B/C > 1$) (6)	Mean nominal cost (7)	Studies / arms
Overall	2.51 [2.15, 2.87]	100%	4.34 [2.94, 5.76]	100%	9.55 [8.11, 11.02]	100%	$961	32 / 43
Productive/lump-sum transfer: with	2.54 [2.15, 2.93]	100%	4.19 [2.86, 5.55]	100%	9.40 [7.86, 10.97]	100%	$961	29 / 36
Productive/lump-sum transfer: without	2.74 [1.60, 3.86]	100%	5.94 [3.10, 8.95]	100%	10.98 [6.63, 15.44]	100%	$567	7 / 7
Consumption support: with	1.95 [1.58, 2.31]	100%	3.38 [2.40, 4.40]	100%	6.89 [5.56, 8.22]	100%	$1,135	25 / 35
Consumption support: without	4.21 [3.33, 5.09]	100%	7.05 [4.14, 9.99]	100%	17.44 [13.48, 21.44]	100%	$369	8 / 8
Coaching: with	2.63 [2.25, 3.02]	100%	4.48 [3.06, 5.93]	100%	10.01 [8.47, 11.57]	100%	$993	30 / 41
Coaching: without	0.75 [0.44, 1.06]	6%	2.28 [0.95, 3.63]	97%	2.90 [1.69, 4.09]	100%	$500	2 / 2
Savings support: with	2.62 [2.23, 3.02]	100%	4.46 [2.99, 5.96]	100%	9.85 [8.29, 11.42]	100%	$988	28 / 39
Savings support: without	1.23 [0.79, 1.67]	85%	2.65 [1.66, 3.68]	100%	5.71 [3.68, 7.78]	100%	$1,102	4 / 4

Notes: Cells report posterior means and 95 percent credible intervals in brackets. Program-weighted B/C first aggregates benefits and costs across arms or evaluations of the same underlying program and then gives each program equal weight. Mean nominal cost is the descriptive mean per household in constant 2020 nominal USD, using reported nominal costs where available and PPP-backconverted nominal equivalents otherwise. Benefit/cost ratios continue to use reported program cost in constant 2020 PPP USD as the denominator. Arms without reported PPP costs are excluded from CBA averages; predicted costs are not substituted. The final column reports studies / treatment arms. With/without comparisons are descriptive, not causal. Linear decay applies 12.5 percent annual dissipation: 50 percent after four years and full decay after eight. Estimated persistence projects benefits over a common 10-year horizon integrating over the full posterior distribution of the slope from the pooled Graduation consumption follow-up model and floors projected consumption effects at zero. Permanent persistence holds the final observed annual consumption effect constant forever, beginning one year after the final observed follow-up. Training is omitted because every eligible CBA arm includes it.

Appendix Tables

Table A.12: Benefit/Cost Ratios by Implementation Era: Graduation and Unconditional Cash

Program type	Implementation era (1)	Linear decay (2)	Estimated persistence (3)	Permanent persistence (4)	Programs / arms or estimates (5)
Graduation	Through 2015	1.36 [0.85, 1.88]	2.89 [1.57, 4.27]	4.35 [2.49, 6.20]	13 / 14
Unconditional cash	Through 2015	2.03 [1.34, 2.74]	2.50 [1.75, 3.48]	9.68 [6.17, 13.24]	28 / 47
Graduation	2016 or later	3.34 [2.85, 3.83]	5.39 [3.80, 7.00]	13.32 [11.24, 15.41]	18 / 29
Unconditional cash	2016 or later	1.61 [0.98, 2.25]	1.93 [1.25, 2.87]	7.95 [4.73, 11.23]	13 / 28

Notes: Cells report posterior means and 95 percent credible intervals in brackets. Earlier programs were implemented through 2015; recent programs were implemented in 2016 or later. Graduation ratios divide consumption benefits by reported full program cost in constant 2020 PPP USD. Arms without reported PPP costs are excluded from Graduation averages; predicted costs are not substituted. Cash ratios divide consumption benefits by the amount transferred plus an administrative-cost allowance equal to 24 percent of the transfer. Linear decay applies 12.5 percent annual dissipation. Estimated persistence projects benefits over a common 10-year horizon, integrating over the full posterior distribution of the corresponding pooled consumption follow-up slope and flooring effects at zero. Permanent persistence holds each program's final observed annual effect constant forever. Comparisons across program type and implementation era are descriptive rather than causal.

Additional robustness checks

Hierarchical models and priors. For reported effect $\widehat{\tau}_i$ and its standard error s_i, the Graduation models use $\widehat{\tau}_i \mid \tau_i \sim N(\tau_i, s_i^2)$ and $\tau_i \sim N(x_i'\beta, \sigma_\tau^2)$. The pooled model replaces $x_i'\beta$ with μ. Define s as the sample standard deviation of the reported effects in the actual estimation sample, after outcome normalization and sample filtering; use $s = 1$ when this standard deviation is undefined or nonpositive. The priors are $\mu \sim N(0, s^2)$, $\beta_k \sim N(0, s^2)$ (including the intercept), and $\sigma_\tau \sim \text{HalfNormal}(0, s^2)$.

The dependence sensitivity adds $u_{g(i)}$ to the latent mean, where $g(i)$ is the randomization, $u_g \sim N(0, \sigma_r^2)$, and $\sigma_r \sim \text{HalfNormal}(0, s^2)$. Baseline and randomization-effect fits use identical observations. This permits shared latent effects while retaining conditionally independent sampling errors.

The cash models scale the reported effects and standard errors by their sample standard deviation. In those scaled units, the unconditional model uses $\mu \sim N(0, 2^2)$ and $\sigma_\theta \sim \text{HalfNormal}(0, 1)$. The cash time model uses $\alpha \sim N(0, 2^2)$, $\beta_t \sim N(0, 1)$, $\log \sigma_{\text{program}} \sim N(-1, 0.7^2)$, and $\log \sigma_\theta \sim N(-1, 0.7^2)$, with standard-normal noncentered latent variables. These differ from the Graduation empirical-scale priors and are reported separately.

Persistence assumptions. The within/between specification writes follow-up time as its mean for arm a, $\bar{t}_a$, and its within-arm deviation $t_{aw} - \bar{t}_a$.

For the ten-year model-based scenario, let $A_a^{(m)}$ be the annualized consumption-effect draw at the first eligible follow-up t_a. With slope draw $b^{(m)}$, projected annual benefits are

$$B_{at}^{(m)} = \max\{0, A_a^{(m)} + b^{(m)}(t - t_a)\}, \qquad t = 1, \ldots, 10, \qquad R_a^{(m)} = \frac{1}{C_a} \sum_{t=1}^{10} \frac{B_{at}^{(m)}}{(1.05)^t}.$$

The signed scenario omits the maximum with zero; the flat scenario sets $b^{(m)} = 0$. All rows retain the same horizon, anchors, costs, programs and weights. The path back-casts and forecasts around the anchor, and the zero floor applies to every evaluated year. Program ratios aggregate benefits and costs across arms before giving each program equal weight.

The permanent-persistence benchmark uses the observed benefit sequence through T_a and continuation beginning at $T_a + 1$. The terminal value at T_a is $\tau_{a,T_a}^{(m)}/r$, with $r = 0.05$ for Graduation and cash. The ten-year projection and linear-decay scenarios use their separately defined benefit paths.

Alternative costs and influence. The cost-gradient panels retain the main-paper specifications and compare reported with predicted PPP costs. Matched panels require both cost measures to be finite and positive and use identical observations; available-sample panels retain the original membership rules. Unit-cost models and leave-one-randomization-out validation document prediction performance. The full PPP cost model supplies predicted costs for the cost-adjusted and predicted-cost analyses. The component-only PPP model supplies fallback costs for the value-for-money analysis when reported PPP costs are unavailable. Both coefficient sets and leave-one-randomization-out prediction checks are reported in Appendix Tables A.18 and A.19.

Appendix Tables

Program-influence checks omit one program at a time from the preferred estimated-persistence calculation, preserving equal weights among the remaining programs.

Table A.13: Dependence robustness

Outcome	Specification	Obs/arms/RCTs	Baseline [95% CrI]	Pr($\beta > 0$)	Randomization RE [95% CrI]	Pr($\beta > 0$)
Consumption	Year trend	52/52/32	1.38 [-1.72, 4.38]	0.817	1.56 [-2.02, 5.13]	0.804
Consumption	Year trend, cost-adjusted	34/34/23	2.94 [0.31, 5.7]	0.983	2.89 [-0.197, 5.98]	0.968
Consumption	Year trend, component-adjusted	52/52/32	4.14 [0.041, 8.15]	0.977	4.63 [-0.126, 9.47]	0.971
Consumption	Year trend, effect per cost	48/48/30	0.113 [-0.0467, 0.296]	0.923	0.145 [-0.0857, 0.405]	0.91
Consumption	Adjusted 2016+ contrast	52/52/32	39.6 [3.53, 73.9]	0.985	42.8 [2.87, 85.1]	0.981
Income	Year trend	52/52/32	1.84 [-0.292, 3.9]	0.958	1.9 [-0.467, 4.28]	0.942
Income	Year trend, cost-adjusted	38/38/25	2.8 [0.718, 4.67]	0.994	2.48 [-0.0561, 4.77]	0.973
Income	Year trend, component-adjusted	52/52/32	3.4 [-0.175, 6.87]	0.969	3.3 [-0.602, 6.96]	0.959
Income	Year trend, effect per cost	50/50/31	0.104 [-0.065, 0.299]	0.898	0.159 [-0.0751, 0.409]	0.917
Income	Adjusted 2016+ contrast	52/52/32	36 [11.1, 61.6]	0.997	32.8 [4.97, 58.8]	0.99
Annual consumption	Pooled timing	82/52/32	47 [-14.8, 106]	0.939	32 [-20.9, 86.2]	0.883

Paired specifications use identical prepared observations. Coefficients are posterior medians; intervals are central 95%. Trends use monthly effects per implementation year (per $100 cost for VFM); era contrasts use monthly effects. Persistence uses annual consumption effects per follow-up year. Shared latent randomization effects do not model sampling-error covariances.

Table A.14: Ten-year persistence assumptions

Scenario	Overall	Earlier	Recent	Recent minus earlier	Pr($\Delta > 0$)
full_posterior__floor_zero	4.35 [2.96, 5.8]	2.9 [1.61, 4.3]	5.41 [3.85, 7.01]	2.51 [1.5, 3.47]	1
full_posterior__signed	4.31 [2.89, 5.76]	2.85 [1.48, 4.28]	5.36 [3.77, 6.96]	2.5 [1.46, 3.55]	1
zero_slope__floor_zero	3.28 [2.79, 3.8]	1.92 [1.34, 2.63]	4.26 [3.54, 4.98]	2.34 [1.36, 3.27]	1
zero_slope__signed	3.22 [2.7, 3.76]	1.87 [1.13, 2.61]	4.2 [3.47, 4.93]	2.34 [1.33, 3.36]	1

Posterior means [central 95% credible intervals]; equal program weights, 5% annual discount rate. All rows use identical programs and draws. First-follow-up anchor, years 1–10 including back-casting; flooring applies to every evaluated year. The signed variants retain negatives. Scenarios are not bounds.

Table A.15: Persistence slopes

Model	Term	Obs/arms/RCTs	Median [95% CrI]	Pr($\beta > 0$)
Baseline regression	Follow-up year	82/52/32	47 [-14.8, 106]	0.939
Randomization intercept	Follow-up year	82/52/32	32 [-20.9, 86.2]	0.883
Baseline regression	Arm mean follow-up	82/52/32	56 [-20.5, 135]	0.922
Baseline regression	Within-arm follow-up	82/52/32	32.4 [-58.8, 124]	0.764

Annual consumption PPP USD per follow-up year, participant households. Within/between estimates are descriptive, not an identified universal forecasting parameter.

Table A.16: Single-program omission sensitivity

Omitted program	Earlier	Recent	Difference	Pr($\Delta > 0$)	Increase of means (%)
None (full sample)	2.9 [1.61, 4.3]	5.41 [3.85, 7.01]	2.51 [1.5, 3.47]	1	86.6
DRC, Angelucci et al. (2023)	2.9 [1.61, 4.3]	5.57 [4.01, 7.13]	2.67 [1.63, 3.66]	1	92.1
Pakistan, Banerjee et al. (2015)	3.05 [1.67, 4.54]	5.41 [3.85, 7.01]	2.36 [1.33, 3.34]	1	77.5
Peru, Banerjee et al. (2015)	3.08 [1.7, 4.58]	5.41 [3.85, 7.01]	2.33 [1.29, 3.31]	1	75.5
Honduras, Banerjee et al. (2015)	3.07 [1.71, 4.54]	5.41 [3.85, 7.01]	2.33 [1.29, 3.32]	1	75.8
Bangladesh, Beam et al. (2025)	2.9 [1.61, 4.3]	5.59 [3.94, 7.29]	2.69 [1.62, 3.72]	1	93
Afghanistan, Bedoya et al. (2019)	2.9 [1.61, 4.3]	5.62 [3.98, 7.31]	2.72 [1.66, 3.74]	1	94.1
Niger, Bossuroy et al. (2022)	2.9 [1.61, 4.3]	5.17 [3.74, 6.62]	2.27 [1.27, 3.24]	1	78.4
Zambia, Botea et al. (2023)	2.9 [1.61, 4.3]	5.55 [3.94, 7.19]	2.65 [1.61, 3.65]	1	91.5
Yemen, Brune et al. (2022)	2.75 [1.45, 4.12]	5.41 [3.85, 7.01]	2.65 [1.76, 3.55]	1	96.3
Kenya, Brudevold-Newman et al. (2024)	3 [1.68, 4.44]	5.41 [3.85, 7.01]	2.41 [1.36, 3.39]	1	80.4
Burkina Faso, adaptive safety nets	2.9 [1.61, 4.3]	5.62 [4.03, 7.24]	2.73 [1.69, 3.73]	1	94.1
Somalia, Leight et al. (2025)	2.9 [1.61, 4.3]	5.6 [3.96, 7.29]	2.71 [1.65, 3.73]	1	93.5
Nicaragua, Macours et al. (2022)	2.98 [1.6, 4.47]	5.41 [3.85, 7.01]	2.42 [1.39, 3.4]	1	81.3
Uganda, Mahmud et al. (2026)	2.9 [1.61, 4.3]	4.64 [3.23, 6.11]	1.75 [0.816, 2.63]	1	60.3
Mauritania, adaptive safety nets	2.9 [1.61, 4.3]	5.6 [3.98, 7.26]	2.7 [1.64, 3.71]	1	93.2
Chad, Premand et al. (2024)	2.9 [1.61, 4.3]	4.86 [3.55, 6.2]	1.96 [1, 2.87]	1	67.7
Bangladesh, Rahman et al. (2021)	2.9 [1.61, 4.3]	4.99 [3.53, 6.53]	2.09 [1.1, 3.04]	1	72.2

Omitted program	Earlier	Recent	Difference	Pr($\Delta > 0$)	Increase of means (%)
Uganda, Sedlmayr et al. (2020)	2.75 [1.59, 4.06]	5.41 [3.85, 7.01]	2.65 [1.58, 3.67]	1	96.2
Senegal, adaptive safety nets	2.9 [1.61, 4.3]	5.39 [3.79, 7.02]	2.5 [1.45, 3.5]	1	86.1
Ethiopia, TUP	3.03 [1.67, 4.52]	5.41 [3.85, 7.01]	2.37 [1.34, 3.36]	1	78.3
Ghana, GUP and savings	3.06 [1.69, 4.55]	5.41 [3.85, 7.01]	2.35 [1.32, 3.34]	1	76.9
India, Bandhan TUP	2.65 [1.44, 4]	5.41 [3.85, 7.01]	2.76 [1.72, 3.78]	1	104
Kenya, REAP full	2.76 [1.56, 4.05]	5.41 [3.85, 7.01]	2.65 [1.58, 3.66]	1	96
Kenya, REAP small	2.83 [1.65, 4.09]	5.41 [3.85, 7.01]	2.58 [1.49, 3.61]	1	91.3
Philippines, DOLE group coaching	2.9 [1.61, 4.3]	5.35 [3.76, 6.98]	2.45 [1.4, 3.44]	1	84.5
Philippines, DOLE individual coaching	2.9 [1.61, 4.3]	5.47 [3.84, 7.14]	2.57 [1.53, 3.58]	1	88.7
Philippines, Graduation program	2.9 [1.61, 4.3]	5.35 [3.76, 6.99]	2.45 [1.42, 3.45]	1	84.7
Uganda, Graduating to Resilience	2.9 [1.61, 4.3]	5.7 [4.06, 7.39]	2.81 [1.74, 3.82]	1	96.9
uganda_gtr_group	2.9 [1.61, 4.3]	5.61 [3.96, 7.3]	2.71 [1.65, 3.74]	1	93.6
uganda_gtr_individual	2.9 [1.61, 4.3]	5.63 [3.99, 7.32]	2.73 [1.67, 3.75]	1	94.3
uganda_wings_no_group	2.65 [1.31, 4.11]	5.41 [3.85, 7.01]	2.75 [1.72, 3.75]	1	104

Preferred estimated-persistence scenario, same equal-program weighting. Each row omits at most one program. No joint exclusions, trimming, or reweighting. Intervals are computed draw by draw; increase of means is a point summary, not a posterior interval.

Table A.17: Alternative reported and predicted costs

Outcome	Cost specification	Sample	Obs/arms	Slope [95% CrI]	Pr($\beta > 0$)
Annual consumption	Reported cost, effect level	available	70/44	11.9 [7.03, 17]	1
Annual consumption	Reported cost, effect per cost	available	70/44	-0.586 [-0.954, -0.266]	0
Annual consumption	Predicted cost, effect level	available	56/34	16.6 [8.85, 24.1]	1
Annual consumption	Predicted cost, effect per cost	available	56/34	-0.419 [-0.888, 0.0147]	0.0293
Annual consumption	Reported cost, effect level	matched	55/33	9.83 [4.46, 15.1]	1
Annual consumption	Reported cost, effect per cost	matched	55/33	-0.589 [-0.955, -0.254]	0.000188
Annual consumption	Predicted cost, effect level	matched	55/33	16.4 [8.61, 23.8]	1
Annual consumption	Predicted cost, effect per cost	matched	55/33	-0.434 [-0.895, -0.00234]	0.0245
Annual income	Reported cost, effect level	available	75/48	-0.209 [-3.74, 3.15]	0.456
Annual income	Reported cost, effect per cost	available	75/48	-0.752 [-1.15, -0.411]	0
Annual income	Predicted cost, effect level	available	62/38	-0.895 [-7.35, 5.66]	0.383
Annual income	Predicted cost, effect per cost	available	62/38	-0.907 [-1.46, -0.388]	0.00075
Annual income	Reported cost, effect level	matched	61/37	-0.78 [-4.54, 3.13]	0.346
Annual income	Reported cost, effect per cost	matched	61/37	-0.807 [-1.19, -0.435]	0
Annual income	Predicted cost, effect level	matched	61/37	-1.48 [-7.55, 5.08]	0.319
Annual income	Predicted cost, effect per cost	matched	61/37	-0.928 [-1.48, -0.442]	0.00025

Cost is in $100 PPP on the horizontal axis. Level effects are annual PPP USD; PerTreatment rows show annual effects per $100 cost. Matched rows retain identical observations with finite positive reported and predicted costs. Available-sample rows reproduce the original sample rules. No cost simulations.

Table A.18: PPP unit-cost models

Model	Term	Estimate	Robust SE	N
full	(Intercept)	6.5	0.717	38
full	clean_imputed_program_scale_log	0.0299	0.0997	38
full	clean_n_prog_gt_treat	0.568	0.287	38
full	clean_pre_existing_infraYes	-1.17	0.226	38
full	clean_qualityLow	-0.489	0.208	38
full	clean_qualityModerate	0.057	0.283	38
full	total_transfer_hundreds	0.0664	0.026	38
full	any_savSupport	0.0773	0.26	38
full	any_coach	-0.0156	0.215	38
full	any_consSupport	0.325	0.272	38
component	(Intercept)	7.12	0.309	49
component	total_transfer_hundreds	0.0956	0.0196	49
component	any_savSupport	-0.252	0.282	49
component	any_coach	-0.521	0.294	49
component	any_consSupport	0.222	0.267	49

Dependent variable: log unit cost in 2020 PPP USD. Full model uses metadata and components; component model uses program components. Heteroskedasticity-robust standard errors.

Table A.19: PPP cost prediction validation

Model	Prediction	N/total	Unpredicted	Log RMSE	R squared	Median absolute error (%)
component	in-sample	49/49	0	0.727	0.376	48.8
component	leave-one-randomization-out	49/49	0	0.844	0.159	61.6
full	in-sample	38/38	0	0.529	0.667	27.6
full	leave-one-randomization-out	38/38	0	0.741	0.347	48.6

Whole randomizations are held out. Predictions unavailable for held-out levels are reported, not silently treated as accurate. This validates costs without replacing the main-paper predictions.

Table A.20: Sampler diagnostics

Model	Specification	Fits	Max R-hat	Min bulk/tail ESS	Divergences	Depth hits	Min E-BFMI	Incomplete/pass
cash	effects	1	1	825/1.52e+03	0	0	0.654	0/1
cash	persistence	1	1	889/983	0	0	0.777	0/1
Randomization intercept	Adjusted 2016+ contrast	2	1.01	619/534	0	0	0.787	0/2
Randomization intercept	Year trend, component-adjusted	2	1	776/777	0	0	0.743	0/2
Randomization intercept	Year trend, cost-adjusted	2	1.01	588/562	0	0	0.698	0/2
Randomization intercept	Year trend, effect per cost	2	1	1.09e+03/1.7e+03	0	0	0.712	0/2
Randomization intercept	Pooled timing	1	1	1.07e+03/1.71e+03	0	0	0.733	0/1
Randomization intercept	Year trend	2	1.01	597/700	0	0	0.687	0/2
Baseline regression	Adjusted 2016+ contrast	2	1	2.62e+03/3.9e+03	0	0	0.697	0/2
Baseline regression	Grouped components	12	1	1.92e+03/2.4e+03	0	0	0.741	0/12
Baseline regression	Predicted cost, effect level	4	1.01	739/1.5e+03	0	0	0.616	0/4
Baseline regression	Predicted cost, effect per cost	4	1.01	807/1.55e+03	0	0	0.545	0/4
Baseline regression	Reported cost, effect level	4	1.01	939/1.6e+03	0	0	0.579	0/4
Baseline regression	Reported cost, effect per cost	4	1.01	645/1.23e+03	0	0	0.492	0/4
Baseline regression	Year trend, component-adjusted	2	1	1.65e+03/2.59e+03	0	0	0.798	0/2
Baseline regression	Year trend, cost-adjusted	2	1	1.5e+03/3.1e+03	0	0	0.621	0/2

Model	Specification	Fits	Max R-hat	Min bulk/tail ESS	Divergences	Depth hits	Min E-BFMI	Incomplete/pass
Baseline regression	implementationYearPublication	2	1	2.24e+03/2.76e+03	0	0	0.875	0/2
Baseline regression	Year trend, effect per cost	2	1	700/1.16e+03	0	0	0.585	0/2
Baseline regression	Pooled timing	1	1	4.48e+03/3.18e+03	0	0	0.981	0/1
Baseline regression	Within/between timing	1	1	2.76e+03/2.51e+03	0	0	0.902	0/1
Baseline regression	Follow-up slope	1	1	732/1.26e+03	0	0	0.6	0/1
Baseline regression	Year trend	4	1	1.22e+03/2.49e+03	0	0	0.887	0/4
Pooled effects	No covariates	15	1	1.67e+03/2.21e+03	0	0	0.615	0/15

Passing requires complete diagnostics for at least four chains, R-hat ¡1.01, bulk/tail ESS ¿400, no divergences/depth hits, E-BFMI ¿=0.3. Missing diagnostics fail validation.

Table A.21: Prior scales by fitted sample

Fit	N	s = SD(reported effects)
Consumption-Year trend-Baseline regression	52	52
Consumption-Year trend-Randomization intercept	52	52
Consumption-Year trend, cost-adjusted-Baseline regression	34	55
Consumption-Year trend, cost-adjusted-Randomization intercept	34	55
Consumption-Year trend, component-adjusted-Baseline regression	52	52
Consumption-Year trend, component-adjusted-Randomization intercept	52	52
Consumption-Year trend, effect per cost-Baseline regression	48	3.85
Consumption-Year trend, effect per cost-Randomization intercept	48	3.85
Consumption-Adjusted 2016+ contrast-Baseline regression	52	52
Consumption-Adjusted 2016+ contrast-Randomization intercept	52	52
Income-Year trend-Baseline regression	52	92.6
Income-Year trend-Randomization intercept	52	92.6
Income-Year trend, cost-adjusted-Baseline regression	38	104
Income-Year trend, cost-adjusted-Randomization intercept	38	104
Income-Year trend, component-adjusted-Baseline regression	52	92.6
Income-Year trend, component-adjusted-Randomization intercept	52	92.6
Income-Year trend, effect per cost-Baseline regression	50	14.1
Income-Year trend, effect per cost-Randomization intercept	50	14.1
Income-Adjusted 2016+ contrast-Baseline regression	52	92.6
Income-Adjusted 2016+ contrast-Randomization intercept	52	92.6

Fit	N	s = SD(reported effects)
Consumption-publication-matched-raw	52	52
Consumption-publication-adjusted	52	52
Income-publication-matched-raw	52	92.6
Income-publication-adjusted	52	92.6
persistence-Baseline regression	82	551
persistence-Randomization intercept	82	551
persistence-within-between	82	551
Annual consumption-Reported cost, effect level-available	70	574
Annual consumption-Reported cost, effect per cost-available	70	0.733
Annual consumption-Predicted cost, effect level-available	56	563
Annual consumption-Predicted cost, effect per cost-available	56	0.364
Annual consumption-Reported cost, effect level-matched	55	560
Annual consumption-Reported cost, effect per cost-matched	55	0.397
Annual consumption-Predicted cost, effect level-matched	55	560
Annual consumption-Predicted cost, effect per cost-matched	55	0.361
Annual income-Reported cost, effect level-available	75	949
Annual income-Reported cost, effect per cost-available	75	1.42
Annual income-Predicted cost, effect level-available	62	1.02e+03
Annual income-Predicted cost, effect per cost-available	62	0.895
Annual income-Reported cost, effect level-matched	61	1.03e+03
Annual income-Reported cost, effect per cost-matched	61	1.55
Annual income-Predicted cost, effect level-matched	61	1.03e+03
Annual income-Predicted cost, effect per cost-matched	61	0.901
heterogeneity-Consumption-baseline	82	45.7
heterogeneity-Consumption-components	82	45.7
heterogeneity-Income-baseline	80	71.6
heterogeneity-Income-components	80	71.6

Fit	N	s = SD(reported effects)
heterogeneity-Food security-baseline	61	0.245
heterogeneity-Food security-components	61	0.245
heterogeneity-Savings-baseline	29	146
heterogeneity-Savings-components	29	146
heterogeneity-Assets-baseline	58	604
heterogeneity-Assets-components	58	604
heterogeneity-Any work-baseline	32	0.12
heterogeneity-Any work-components	32	0.12
heterogeneity-Self-employment-baseline	47	0.0929
heterogeneity-Self-employment-components	47	0.0929
heterogeneity-Mental health-baseline	48	0.115
heterogeneity-Mental health-components	48	0.115
heterogeneity-Life satisfaction-baseline	21	0.227
heterogeneity-Life satisfaction-components	21	0.227
heterogeneity-Aspirations-baseline	24	0.0872
heterogeneity-Aspirations-components	24	0.0872
heterogeneity-Agency-baseline	32	0.136
heterogeneity-Agency-components	32	0.136
heterogeneity-Empowerment-baseline	22	1.1
heterogeneity-Empowerment-components	22	1.1
savings-full	30	144
savings-excluding-ethiopia	29	46.2
cba-effects	83	45.6
cba-slope	82	551

For these Graduation hierarchical models, mean/intercept and regression coefficients have Normal(0,s) priors and heterogeneity SDs HalfNormal(0,s). Randomization SD also has HalfNormal(0,s). Scales are recomputed after sample filtering and outcome normalization; fallback s=1 if undefined or nonpositive. Cash model priors are separately specified in the methods source, not by this scale table.

Table A.22: Publication status by implementation era

Outcome	Implementation era	Publication status	Arms	Randomizations
Consumption	Earlier (through 2015)	Published	16	14
Consumption	Earlier (through 2015)	Unpublished	1	1
Consumption	Earlier (through 2015)	Unknown	0	0
Consumption	Recent (2016+)	Published	6	4
Consumption	Recent (2016+)	Unpublished	29	13
Consumption	Recent (2016+)	Unknown	0	0
Income	Earlier (through 2015)	Published	18	15
Income	Earlier (through 2015)	Unpublished	1	1
Income	Earlier (through 2015)	Unknown	0	0
Income	Recent (2016+)	Published	6	4
Income	Recent (2016+)	Unpublished	27	12
Income	Recent (2016+)	Unknown	0	0

Counts use the full raw implementation-year first-follow-up sample, separately by outcome. Unknown publication status is shown explicitly; it is excluded from the matched regression panel. Randomization counts are distinct within each cell and need not add across cells. No main-text interpretation is changed.

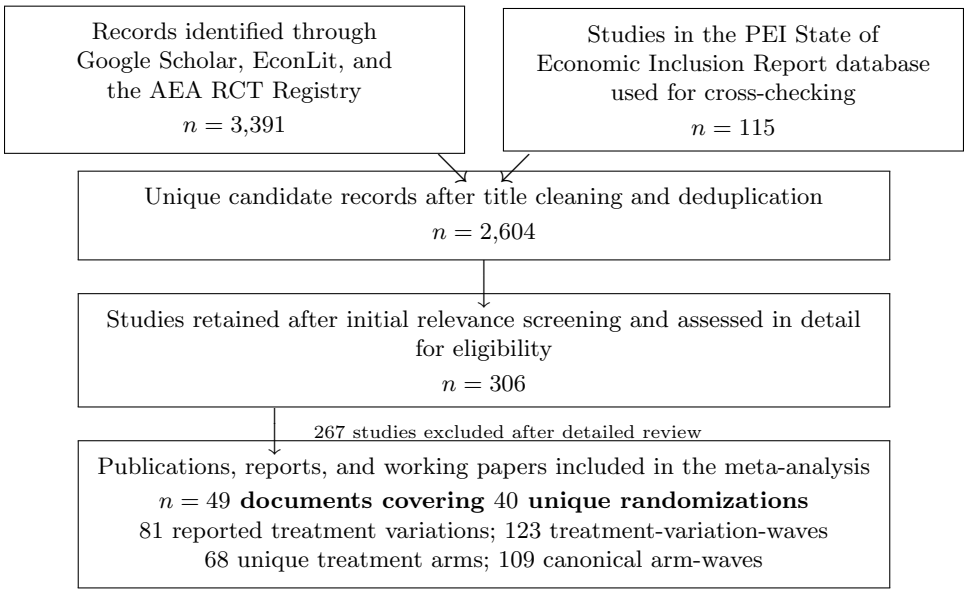
Table A.23: Implementation-year trend and publication status

Outcome	Specification	Obs/arms/RCTs	Year slope [95% CrI]	Pr($\beta > 0$)
Consumption	Raw, matched sample	52/52/32	1.38 [-1.72, 4.38]	0.817
Consumption	Publication-adjusted	52/52/32	1.04 [-3.24, 5.46]	0.684
Income	Raw, matched sample	52/52/32	1.84 [-0.292, 3.9]	0.958
Income	Publication-adjusted	52/52/32	2.47 [-0.736, 5.69]	0.938

Posterior median monthly PPP USD per implementation year. Raw and publication-adjusted fits use the same complete-case first-follow-up sample. The full-sample raw trend is in R1. Publication status and era can overlap strongly; this is a descriptive sensitivity check, not identification of publication bias. Main-paper specifications are unchanged.

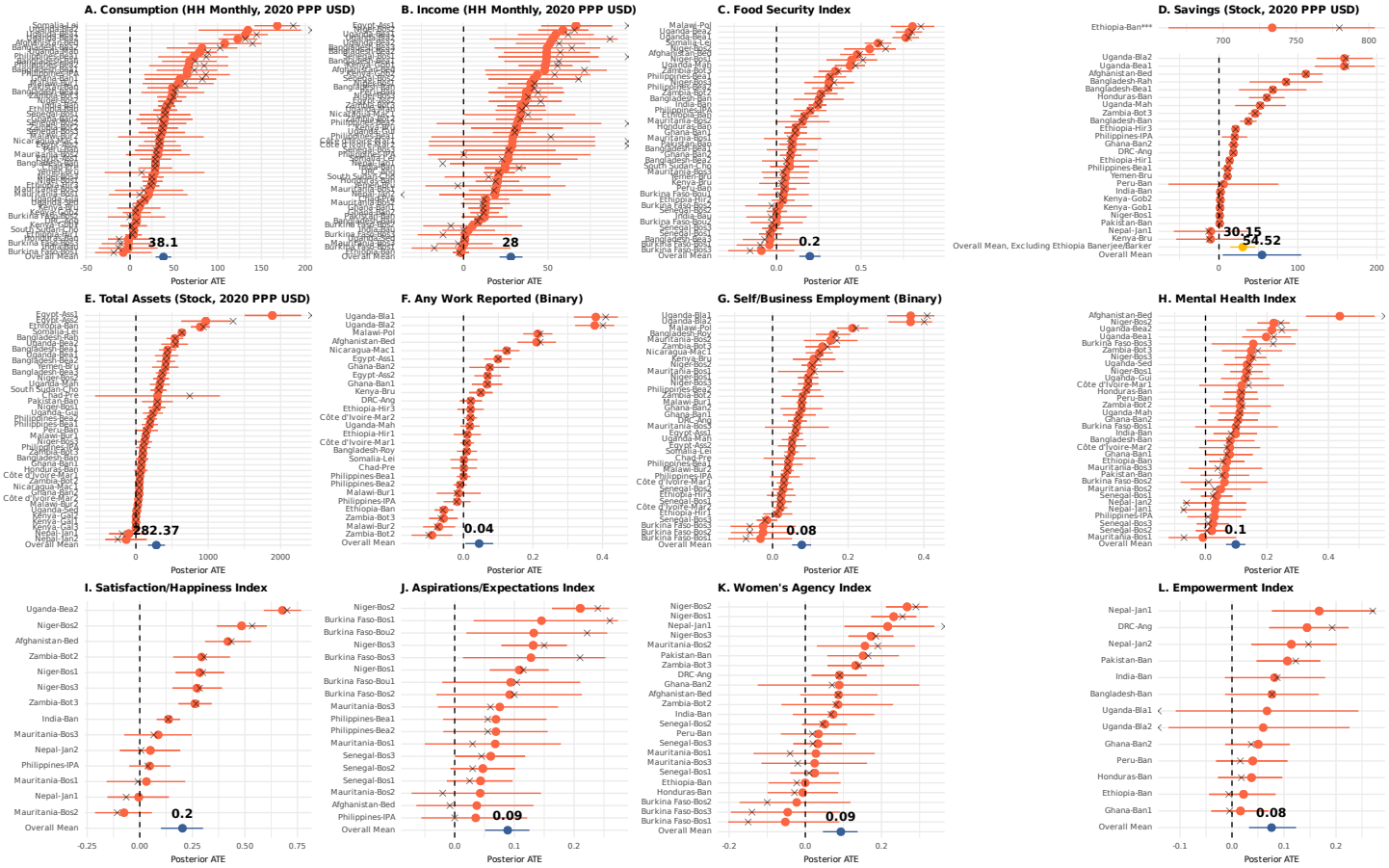
Appendix Figures

Figure A.1: Study search and screening flow



Notes: The PEI database was used to cross-check the database search. Detailed eligibility review applied the study- and treatment-arm-level criteria described in the text. Treatment variations are study-publication arm records; unique treatment arms consolidate repeated reports of the same intervention lineage across publications and follow-ups.

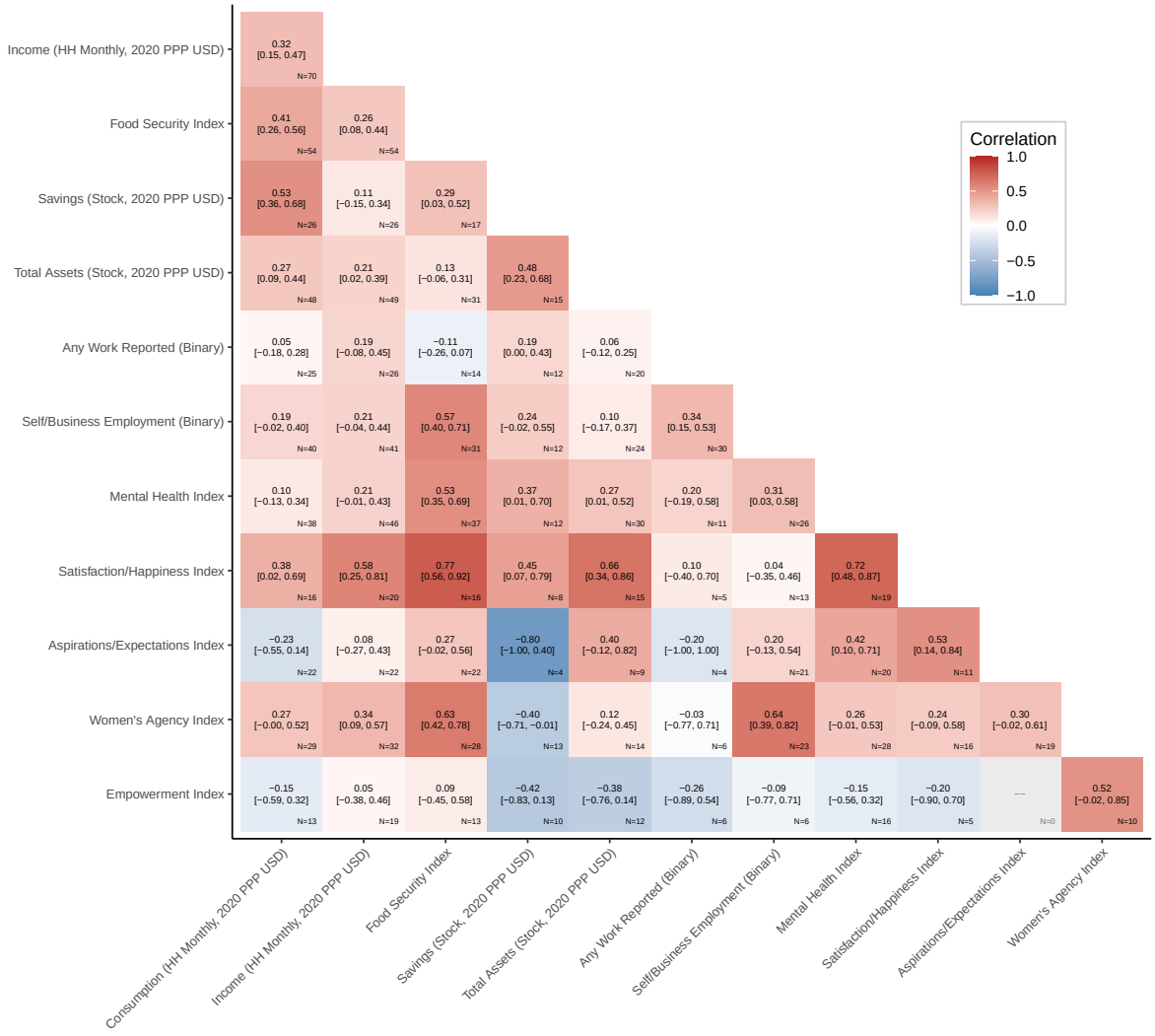
Figure A.2: Forest Plots for All Main Outcomes



Notes: Panels report (A) Consumption, (B) Income, (C) Food Security, (D) Savings, (E) Total Assets, (F) Any Work, (G) Self/Business Employment, (H) Mental Health, (I) Satisfaction/Happiness, (J) Aspirations/Expectations, (K) Women's Agency, and (L) Empowerment. Points and horizontal lines report posterior medians and 95 percent credible intervals for treatment arms, identified using canonical country–author labels. The dashed vertical line marks zero; each panel's "Overall Mean" row reports the pooled posterior result. Horizontal scales are tailored independently and should not be compared across panels. Panel D also reports the pooled savings mean excluding the large Ethiopia estimates from Banerjee et al. (2015a).

Appendix Figures

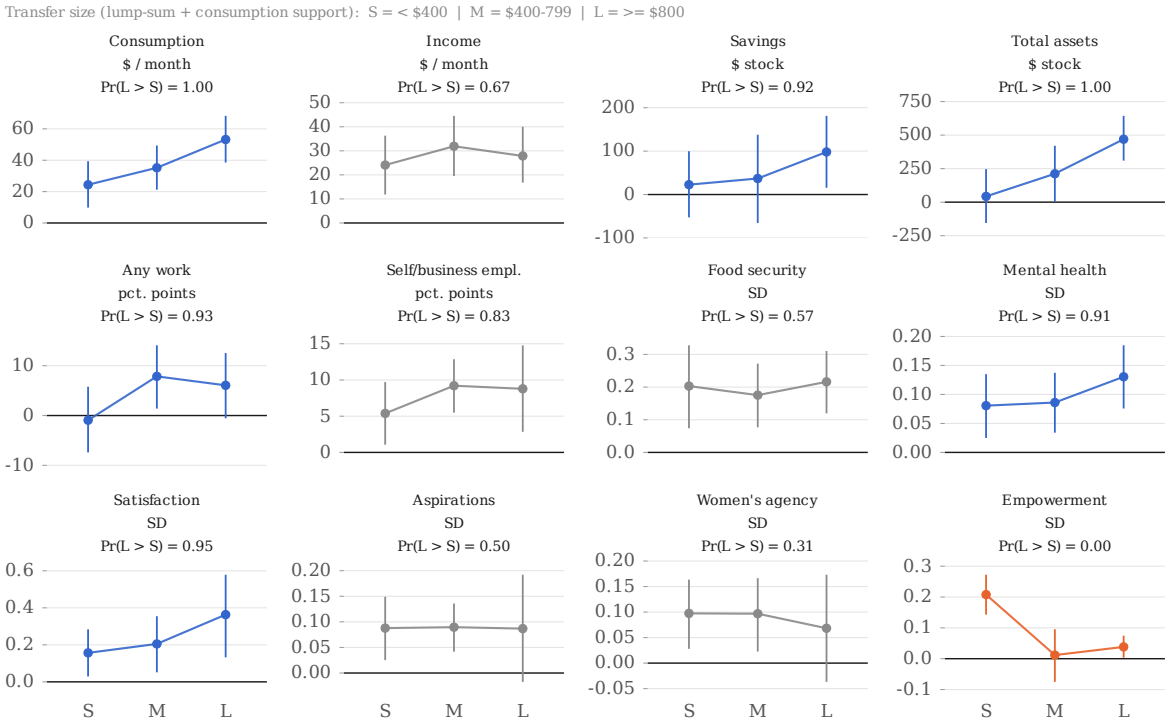
Figure A.3: Spearman Rank Correlations Across Treatment Effects



Notes: The figure reports posterior median pairwise Spearman rank correlations in treatment effects across outcomes, computed across matched treatment-arm waves within each posterior draw. Brackets report 95 percent credible intervals. Sample sizes are pair-specific because outcomes are not observed for every arm-wave. A black cell with "–" indicates that the correlation is not estimable because fewer than four overlapping arm-wave observations are available.

Appendix Figures

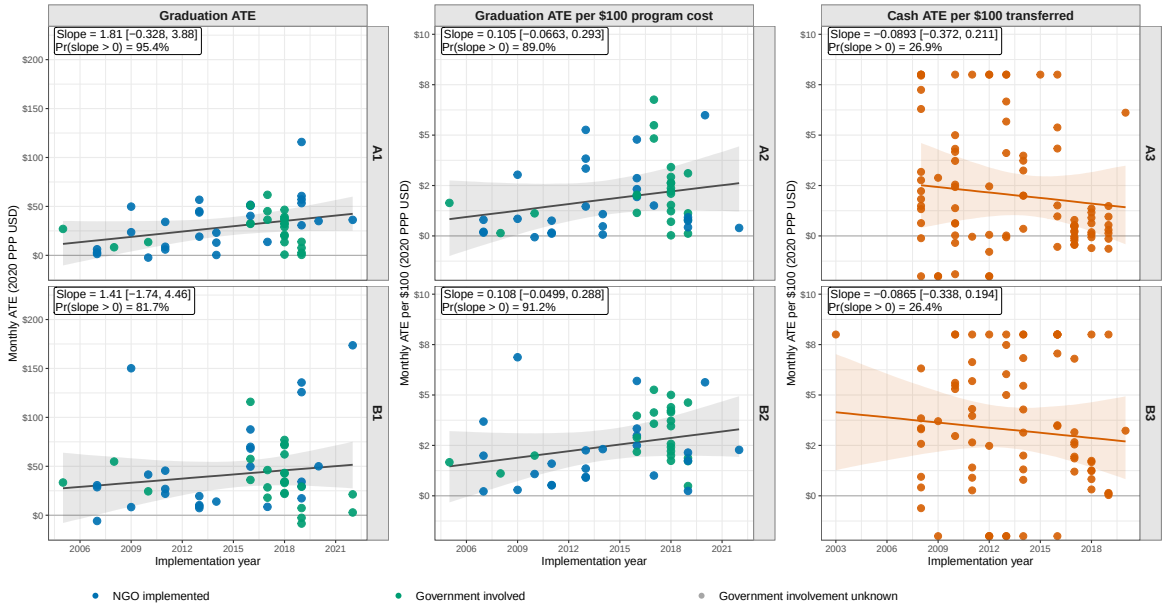
Figure A.4: Treatment Effects by Transfer Size



Notes: The figure reports adjusted treatment effects for programs with no or small transfers (less than $400), medium transfers ($400–$799), and large transfers ($800 or more), measured in 2020 PPP USD and combining the lump-sum productive transfer with consumption support. Points and whiskers report posterior medians and 95 percent credible intervals. Each panel reports the posterior probability that the effect for large transfers exceeds the effect for no or small transfers. Blue indicates a probability of at least 0.9, orange a probability of at most 0.1, and gray an intermediate probability. Vertical scales vary across panels.

Appendix Figures

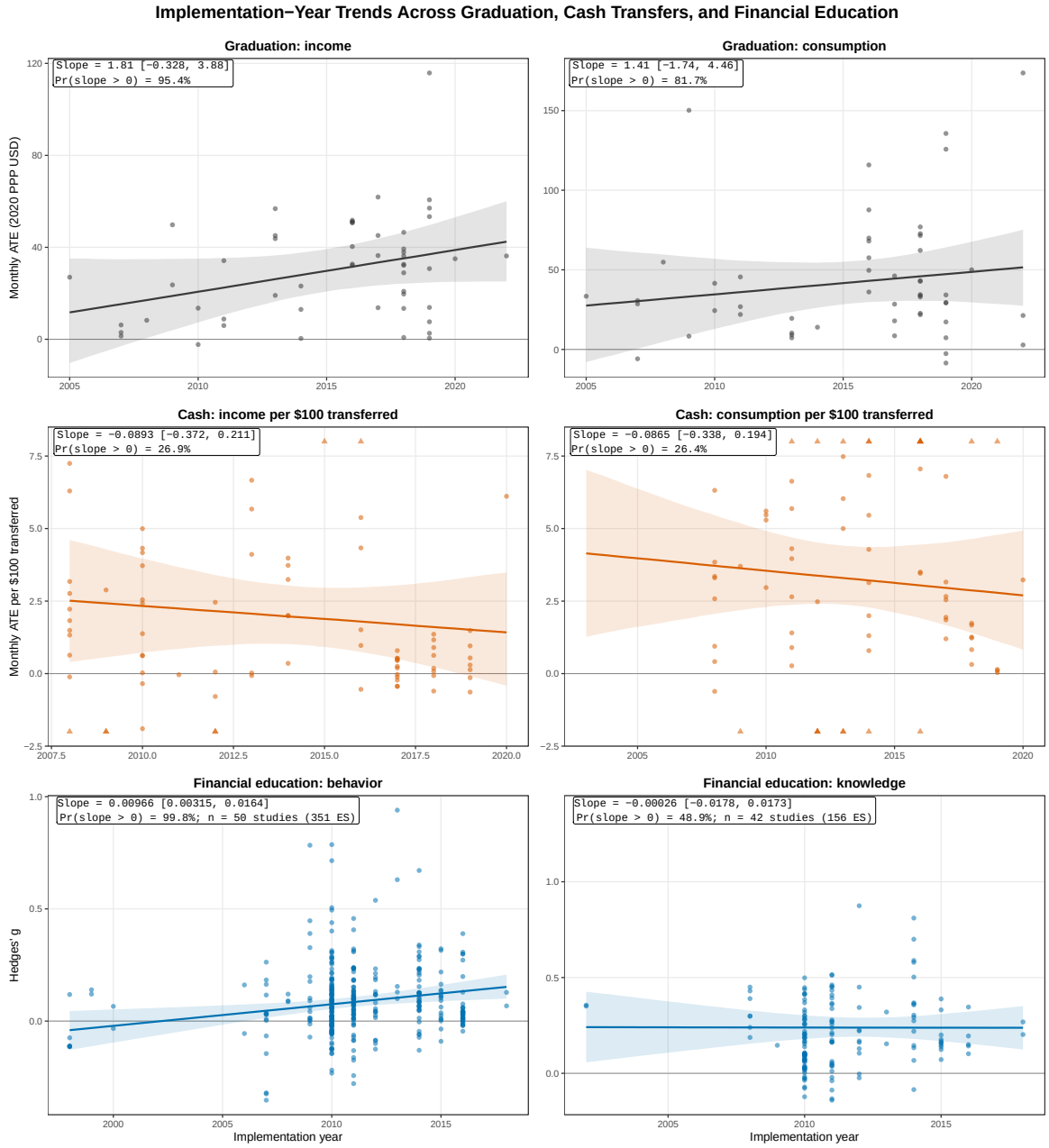
Figure A.5: Treatment Effects and Program Implementation Year, by Implementer Type



Notes: Panels show implementation-year trends in Graduation effects, Graduation effects per unit of program cost, and cash-transfer effects per unit transferred, separately for consumption and income. Blue denotes NGO implementation, green denotes government involvement, and gray denotes unknown government involvement. Orange denotes cash-transfer estimates. The cost- and component-adjusted Graduation specifications appear in Figure 5.

Appendix Figures

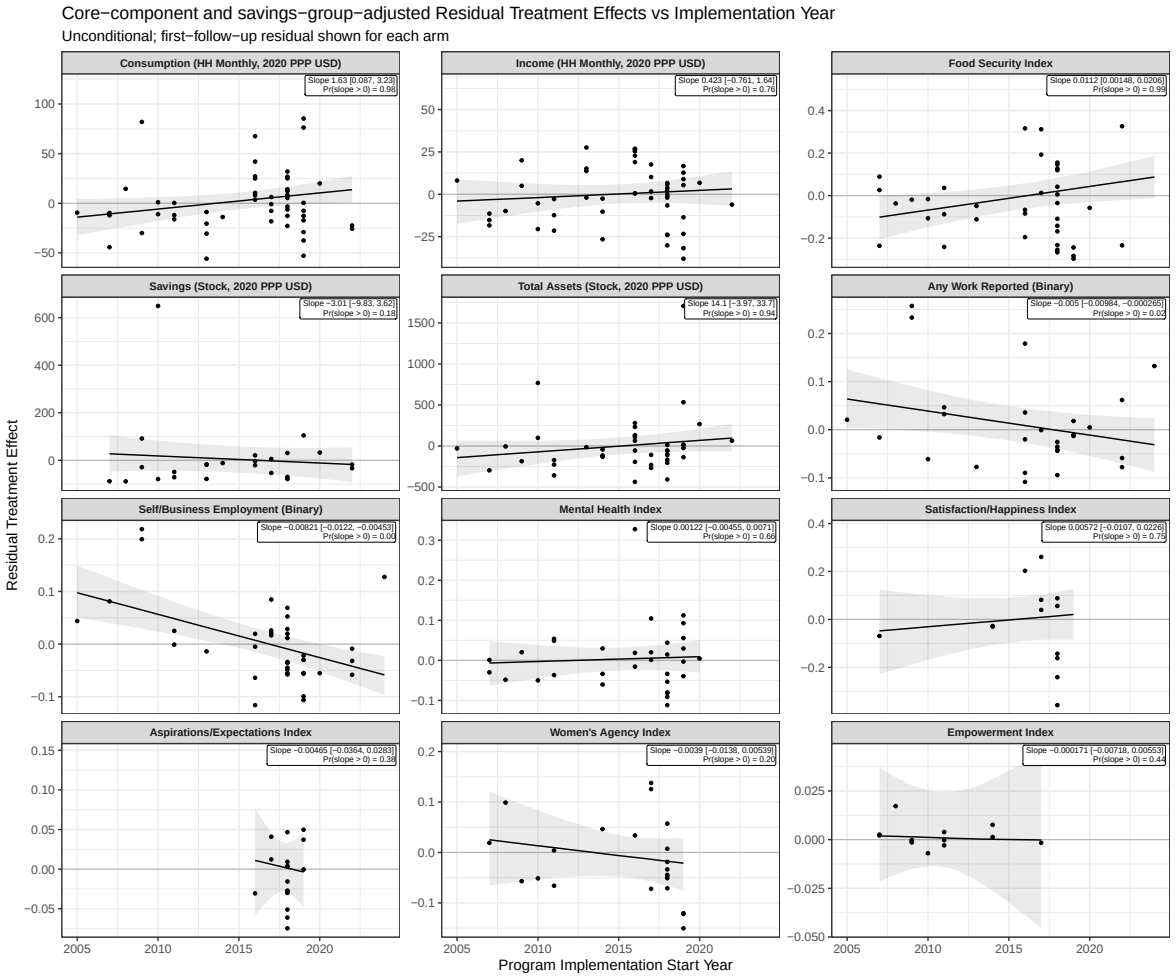
Figure A.6: Implementation-Year Trends Across Graduation, Cash Transfers, and Financial Education



Notes: The top row reports Graduation income and consumption effects in monthly 2020 PPP USD. The middle row reports cash-transfer income and consumption effects in 2020 PPP USD per $100 transferred; baseline year proxies implementation year, and estimates outside the displayed range are retained in estimation and plotted at the relevant boundary as triangles. The bottom row reports financial-education effects on financial behavior and knowledge in Hedges' g, using the replication data assembled by Kaiser et al. (2022a). Graduation specifications use the first follow-up for each treatment arm. Financial-education specifications use the earliest follow-up within each study and outcome, retain multiple effect sizes nested within studies, and control for standardized log weeks between treatment and measurement; fitted lines are evaluated at the sample mean of logged follow-up timing. Lines and shaded regions report posterior medians and 95 percent credible intervals. Panel annotations report posterior median annual slopes, central 95 percent credible intervals in brackets, and posterior probabilities of a positive slope.

Appendix Figures

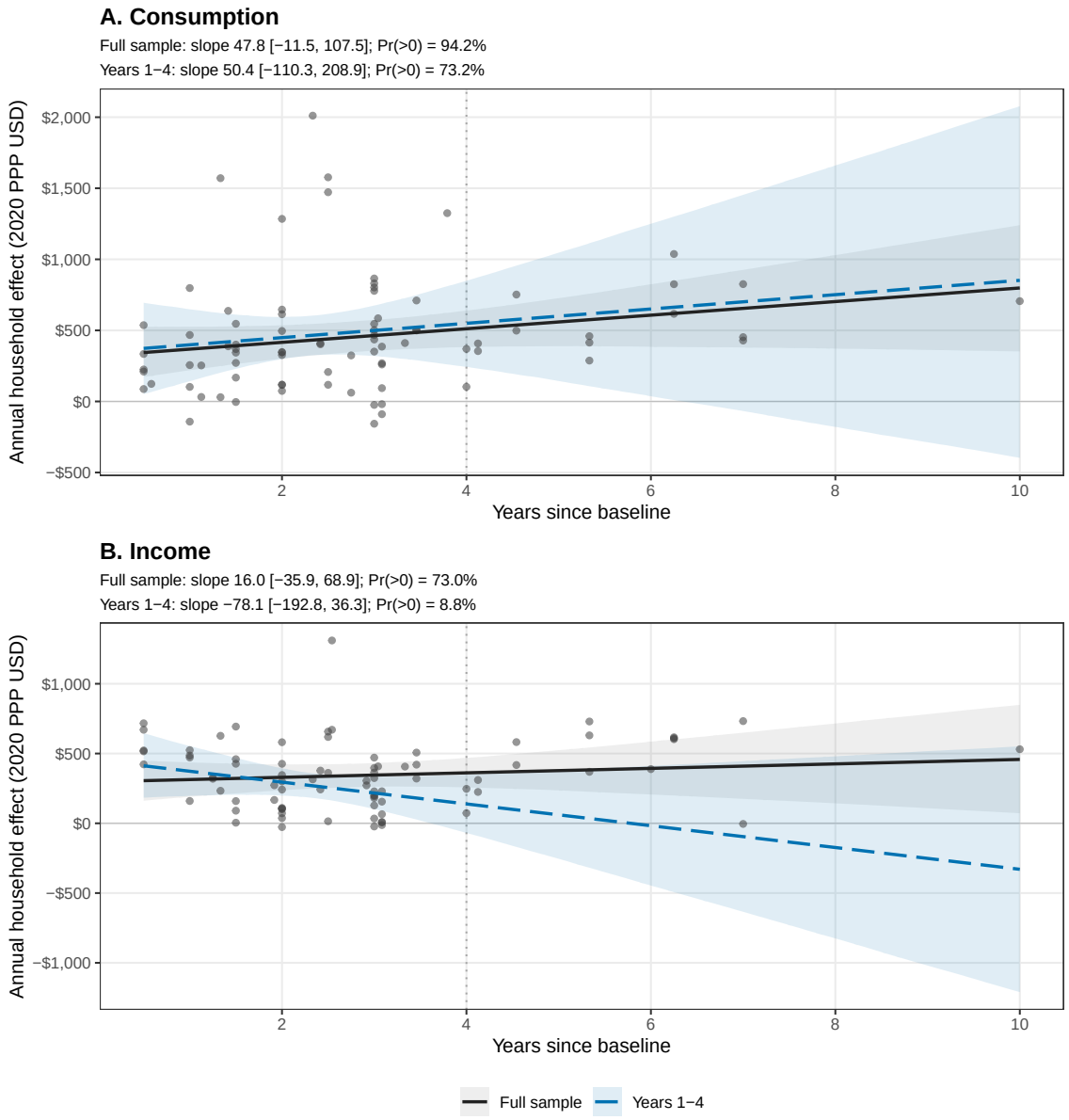
Figure A.7: Component-Adjusted Residual Treatment Effects and Implementation Year: Main Outcomes



Notes: This figure extends the component-adjusted specification in row 4 of Figure 5, to every main outcome defined in the analysis. Panels use outcome-specific vertical scales.

Appendix Figures

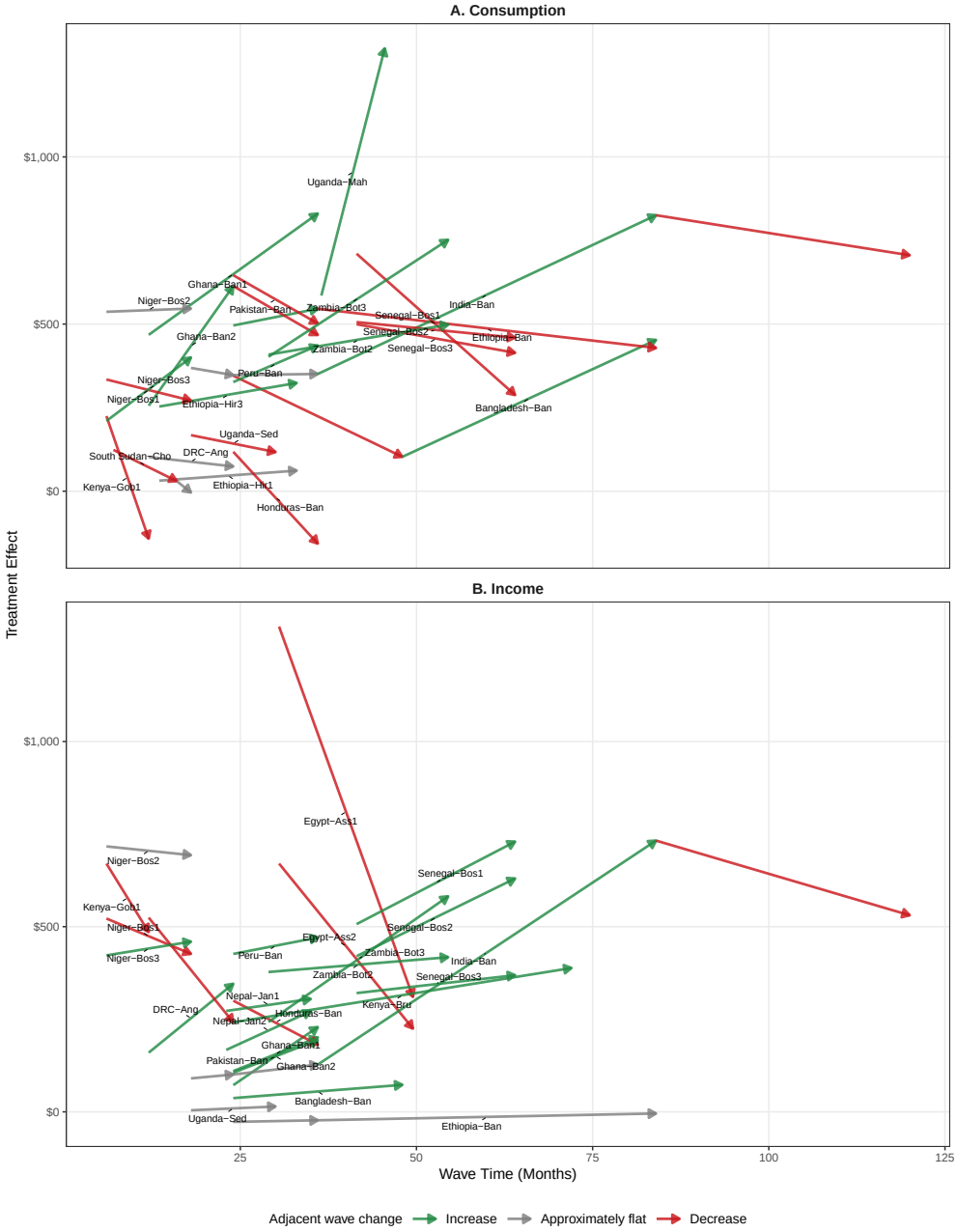
Figure A.8: Persistence using Full Sample or Short-Run Estimates



Notes: Solid black lines use all eligible follow-ups (six months to ten years after baseline); dashed blue lines use only follow-ups from one to four years after baseline. The restricted lines outside years 1–4 are extrapolations. Lines combine the posterior-median intercept and slope; shaded regions are pointwise 95 percent credible intervals for the regression mean. Dots show the full-sample model's shrunk arm-wave effects. The vertical dotted line marks year 4. Effects are annual household consumption (Panel A) and income (Panel B), in 2020 PPP USD. The restricted fits retain 62 observations for each outcome, compared with 82 for consumption and 81 for income in the full sample. The benefit-cost sensitivity uses the restricted consumption slope distribution with the existing 10-year horizon, 5 percent discount rate, and zero floor for annual projected benefits.

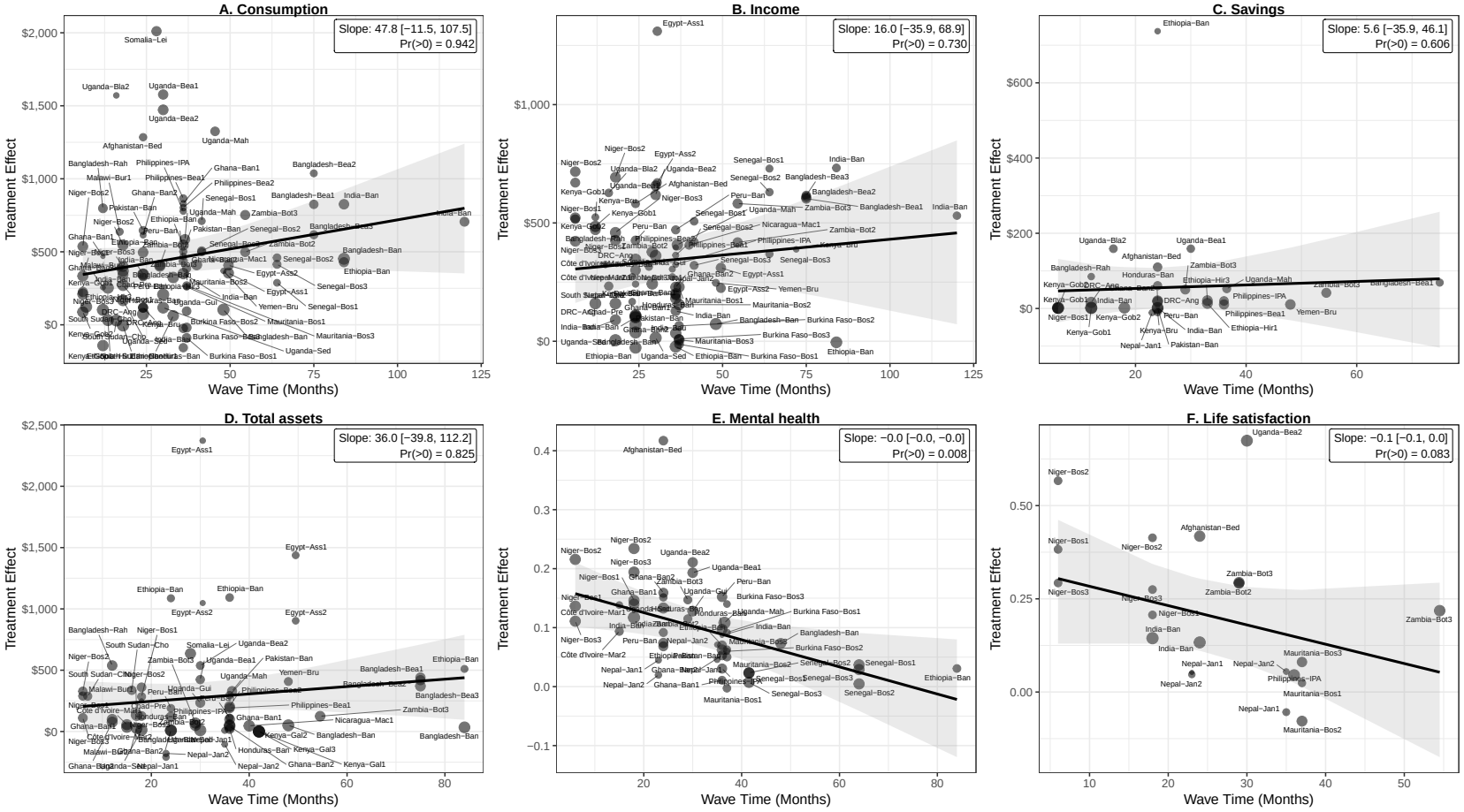
Appendix Figures

Figure A.9: Within-Arm Changes in Consumption and Income Treatment Effects



Notes: The upper and lower panels report annual household consumption and income treatment effects, respectively, in 2020 PPP USD. Each arrow connects two consecutive observed follow-up waves within a treatment arm, with wave time in months on the horizontal axis. Green, gray, and red arrows indicate increases, approximately flat changes, and decreases, respectively. A change is approximately flat when its absolute value is no more than 10 percent of the outcome-specific median absolute starting treatment effect across adjacent-wave pairs. Program labels use the canonical country–author figure codes.

Figure A.10: Persistence of Treatment Effects: Main Outcomes



Notes: Panels A–F report observed study-arm follow-up effects for consumption, income, savings, total assets, mental health, and life satisfaction. Larger points denote more precise estimates within each outcome. Lines report the posterior relationship with wave time estimated from all available follow-ups, and shaded regions are 95 percent credible intervals; panel subtitles report annual slopes, their 95 percent credible intervals, and posterior probabilities that slopes are positive. Panels A and B repeat the main-text specifications for reference. The pooled follow-up-time relationships for mental health and life satisfaction slope downward. These are descriptive associations across the available observations, rather than identified within-program rates of decay.

Data Appendix

Data Appendix Table 1: Studies Included in the Meta-analysis

Country	Study	Implementation Year (1)	Arms (2)	Waves (3)	Arm-waves (4)	Months since program start (5)
Afghanistan	Bedoya et al. (2019)	2016	1	1	1	24
Afghanistan	Gibbs et al. (2020)	2016	1	1	1	22
BRAC Bangladesh TUP 2007						
Bangladesh	Bandiera et al. (2017)	2007	1	3	3	24, 48, 84
Bangladesh	Ahmed et al. (2026)	2007	1	1	1	204
Bangladesh	Beam et al. (2025)	2016	3	1	3	75
Bangladesh	Rahman et al. (2021)	2016	1	1	1	12
Bangladesh	Roy et al. (2015)	2007	1	1	1	60
Burkina Faso	Bossuroy et al. (2026)	2019	3	1	3	37
Burkina Faso	Bouguen et al. (2026)	2018	2	2	4	12, 24
Chad	Premand et al. (2024)	2019	1	1	1	18
Côte d'Ivoire	Marguerie et al. (2023)	2016	2	1	2	15
Congo, Dem. Rep.	Angelucci et al. (2023)	2017	1	2	2	12, 24
Egypt, Arab Rep.	Assaad et al. (2026)	2019	2	2	4	31, 50
Ethiopia TUP						
Ethiopia	Banerjee et al. (2015)	2010	1	2	2	24, 36
Ethiopia	Barker et al. (2024)	2010	1	2	2	36, 84
Ethiopia	Hirvonen et al. (2025)	2022	3	2	5	14, 33
Ghana GUP with savings						
Ghana	Banerjee et al. (2015)	2011	1	2	2	24, 36
Ghana	Banerjee et al. (2025)	2011	2	3	6	12, 24, 36
Honduras	Banerjee et al. (2015)	2009	1	2	2	24, 36
Bandhan TUP/THP						
India	Banerjee et al. (2011)	2007	1	1	1	18
India	Banerjee et al. (2015)	2007	1	3	3	24, 36, 84
India	Banerjee et al. (2021)	2007	1	4	4	18, 36, 84, 120
India	Bauchet et al. (2015)	2007	1	1	1	36
BOMA REAP						
Kenya	Gobin et al. (2016)	2013	2	3	3	6, 12
Kenya	Gobin et al. (2017)	2013	2	3	5	6, 12
Kenya	Brudevold-Newman et al. (2024)	2013	1	3	3	12, 24, 72
Kenya	Gallardo et al. (2021)	2018	3	1	3	42
Malawi	Burchi et al. (2018)	2016	2	1	2	17
Malawi	Poll et al. (2025)	2024	1	1	1	12
Mauritania	Bossuroy et al. (2026)	2018	3	1	3	37
Nepal	Janzen et al. (2021)	2014	2	2	4	23, 35
Nicaragua	Macours et al. (2022)	2005	1	1	1	40
Niger	Bossuroy et al. (2022)	2017	3	2	6	6, 18
Pakistan	Banerjee et al. (2015)	2008	1	2	2	24, 36
Peru	Banerjee et al. (2015)	2011	1	2	2	24, 36
ADB–BRAC–DOLE Graduation of the Ultra-Poor program						
Philippines	IPA (2022)	2018	3	1	3	36
Philippines	Beam et al. (2025)	2018	2	1	2	36
Senegal	Bossuroy et al. (2026)	2018	3	2	6	42, 64
Somalia	Leight et al. (2025)	2022	1	1	1	28
South Sudan	Chowdhury et al. (2017)	2013	1	3	3	7, 16, 18
Graduating to Resilience						
Uganda	Guilhin et al. (2022)	2019	3	1	3	30
Uganda	Beam et al. (2025)	2019	2	1	2	30
Uganda	Mahmud et al. (2026)	2020	1	2	2	37, 46
Uganda	Sedlmayr et al. (2020)	2014	1	3	3	18, 30
WINGS Uganda						
Uganda	Blattman et al. (2016)	2009	2	1	2	16
Uganda	Malaeb et al. (2023)	2009	2	1	2	16
Uganda	Wang et al. (2021)	2012	2	1	2	24
Yemen, Rep.	Brune et al. (2022)	2010	1	1	1	48
Zambia	Botea et al. (2023)	2018	2	2	4	29, 55

Notes: The 49 study records, drawn from 41 publications, reports, and working papers, cover 40 unique randomizations. Indented rows under a shared heading are documents reporting the same randomization, including later follow-ups. Implementation year is the year program delivery began. Arms are distinct eligible treatment arms and waves are distinct follow-up survey rounds reported in each source. Arm-waves count the distinct treatment-arm-by-follow-up combinations in that source. Column (5) reports when impacts were measured, in months since program start or baseline (the timing used in the persistence analysis), rounded to the nearest month; multi-wave sources list each follow-up. Counts can overlap across rows within a study group and should not be summed.

Data Appendix Table 2: Program Components by Treatment Arm

Country	Figure Code	Study	Treatment Arm	Implementation Year	Transfer	Cons. Support	Savings	Coaching	Training	Duration (months)
				(1)	(2)	(3)	(4)	(5)	(6)	
Afghanistan	Afghanistan-Bed	Bedoya et al. (2019)	Targeting the Ultra Poor	2016	2,031	53	Yes	Yes	Yes	12.0
Afghanistan	Afghanistan-Gib	Gibbs et al. (2020)	Empowerment program	2016	0	553	Yes	No	Yes	12.0
Bangladesh	Bangladesh-Ban	Bandiera et al. (2017)	Targeting the Ultra Poor	2007	714	460	Yes	Yes	Yes	24.0
Bangladesh	Bangladesh-Bea1	Beam et al. (2025)	Group coaching	2016	1,180	637	Yes	Yes	Yes	18.0
Bangladesh	Bangladesh-Bea2	Beam et al. (2025)	Individual coaching	2016	1,180	637	Yes	Yes	Yes	18.0
Bangladesh	Bangladesh-Bea3	Beam et al. (2025)	Light individual coaching	2016	1,180	637	Yes	Yes	Yes	18.0
Bangladesh	Bangladesh-Rah	Rahman et al. (2021)	Hybrid UPG	2016	0	638	Yes	Yes	Yes	18.0
Bangladesh	Bangladesh-Roy	Roy et al. (2015)	Treatment	2007	509	561	Yes	Yes	Yes	24.0
Burkina Faso	Burkina Faso-Bos2	Bossuroy et al. (2026)	Graduation	2019	470	0	Yes	Yes	Yes	18.0
Burkina Faso	Burkina Faso-Bos3	Bossuroy et al. (2026)	Psychosocial	2019	0	0	Yes	Yes	Yes	18.0
Burkina Faso	Burkina Faso-Bos1	Bossuroy et al. (2026)	Full	2019	470	0	Yes	Yes	Yes	18.0
Burkina Faso	Burkina Faso-Bou1	Bouguen et al. (2026)	T2	2018	354	720	No	No	Yes	–
Burkina Faso	Burkina Faso-Bou2	Bouguen et al. (2026)	T3	2018	354	–	No	No	Yes	–
Chad	Chad-Pre	Premand et al. (2024)	Treatment	2019	248	0	Yes	Yes	Yes	3.5
Côte d'Ivoire	Côte d'Ivoire-Mar1	Marguerie et al. (2023)	Cash grant	2016	513	0	No	Yes	Yes	9.0

Data Appendix Table 2: Program Components by Treatment Arm *(continued)*

Country	Figure Code	Study	Treatment Arm	Implementation Year (1)	Transfer (2)	Cons. Support (3)	Savings (4)	Coaching (5)	Training (6)	Duration (months)
Côte d'Ivoire	Côte d'Ivoire-Mar2	Marguerie et al. (2023)	Cash grant + repayment	2016	258	0	No	Yes	Yes	9.0
Congo, Dem. Rep.	DRC-Ang	Angelucci et al. (2023)	Stronger Nations Stronger Women	2017	0	316	Yes	No	Yes	12.0
Egypt, Arab. Rep.	Egypt-Ass1	Assaad et al. (2026)	Full-Transfer Unrestricted + Full-Transfer Restricted	2019	2,837	825	Yes	Yes	Yes	–
Egypt, Arab. Rep.	Egypt-Ass2	Assaad et al. (2026)	Half-Transfer	2019	1,418	619	Yes	Yes	Yes	–
Ethiopia	Ethiopia-Ban	Barker et al. (2024)	–	2010	1,352	0	Yes	Yes	Yes	24.0
Ethiopia	Ethiopia-Hir1	Hirvonen et al. (2025)	T2 (PSNP + SPIR II graduation programming + Nurturing Care Groups)	2022	214	0	Yes	Yes	Yes	–
Ethiopia	Ethiopia-Hir2	Hirvonen et al. (2025)	T2 (PSNP + SPIR II graduation programming + Nurturing Care Groups) + T3 (PSNP + SPIR II graduation programming + NCGs + maternal cash transfers)	2022	214	525	Yes	Yes	Yes	–
Ethiopia	Ethiopia-Hir3	Hirvonen et al. (2025)	T3 (PSNP + SPIR II graduation programming + NCGs + maternal cash transfers)	2022	214	1,049	Yes	Yes	Yes	–
Ghana	Ghana-Ban2	Banerjee et al. (2025)	Graduating from Ultra Poverty with savings	2011	498	498	Yes	Yes	Yes	24.0

Data Appendix Table 2: Program Components by Treatment Arm *(continued)*

Country	Figure Code	Study	Treatment Arm	Implementation Year (1)	Transfer (2)	Cons. Support (3)	Savings (4)	Coaching (5)	Training (6)	Duration (months)
Ghana	Ghana-Ban1	Banerjee et al. (2025)	Graduating from Ultra Poverty without savings	2011	498	498	No	Yes	Yes	24.0
Honduras	Honduras-Ban	Banerjee et al. (2015)	–	2009	592	146	Yes	Yes	Yes	24.0
India	India-Ban	Banerjee et al. (2021)	TUP program	2007	481	91	Yes	Yes	Yes	18.0
India	India-Bau	Bauchet et al. (2015)	Treatment	2007	447	0	Yes	Yes	Yes	18.0
Kenya	Kenya-Bru	Brudevold-Newman et al. (2024)	Franchise treatment	2013	795	0	No	Yes	Yes	2.3
Kenya	Kenya-Gal1	Gallardo et al. (2021)	Large Grant Village Enterprise Program	2018	485	0	Yes	Yes	Yes	10.5
Kenya	Kenya-Gal2	Gallardo et al. (2021)	Small Grant Village Enterprise Program	2018	149	0	Yes	Yes	Yes	10.5
Kenya	Kenya-Gal3	Gallardo et al. (2021)	Village Enterprise Program	2018	149	0	Yes	Yes	Yes	10.5
Kenya	Kenya-Gob1	Gobin et al. (2017)	Group A	2013	382	0	Yes	Yes	Yes	24.0
Kenya	Kenya-Gob2	Gobin et al. (2017)	Group B	2013	382	0	Yes	Yes	Yes	24.0
Malawi	Malawi-Bur1	Burchi et al. (2018)	Lump sum plus training	2016	247	163	Yes	Yes	Yes	–
Malawi	Malawi-Bur2	Burchi et al. (2018)	Training only	2016	0	0	Yes	Yes	Yes	–
Malawi	Malawi-Pol	Poll et al. (2025)	Treatment	2024	921	371	Yes	Yes	Yes	12.0
Mauritania	Mauritania-Bos1	Bossuroy et al. (2026)	Capital	2018	687	0	Yes	Yes	Yes	–

Data Appendix Table 2: Program Components by Treatment Arm *(continued)*

Country	Figure Code	Study	Treatment Arm	Implementation Year	Transfer	Cons. Support	Savings	Coaching	Training	Duration (months)
				(1)	(2)	(3)	(4)	(5)	(6)	
Mauritania	Mauritania-Bos3	Bossuroy et al. (2026)	Psychosocial	2018	0	0	Yes	Yes	Yes	–
Mauritania	Mauritania-Bos2	Bossuroy et al. (2026)	Full	2018	687	0	Yes	Yes	Yes	–
Nepal	Nepal-Jan1	Janzen et al. (2021)	Full treatment	2014	181	0	Yes	No	Yes	–
Nepal	Nepal-Jan2	Janzen et al. (2021)	No values based training	2014	181	0	Yes	No	Yes	–
Nicaragua	Nicaragua-Mac1	Macours et al. (2022)	CCT and Grant	2005	254	305	No	No	Yes	6.0
Niger	Niger-Bos1	Bossuroy et al. (2022)	Capital arm	2017	350	398	Yes	Yes	Yes	10.5
Niger	Niger-Bos2	Bossuroy et al. (2022)	Full arm	2017	350	398	Yes	Yes	Yes	10.5
Niger	Niger-Bos3	Bossuroy et al. (2022)	Psychosocial arm	2017	0	398	Yes	Yes	Yes	10.5
Pakistan	Pakistan-Ban	Banerjee et al. (2015)	–	2008	1,151	77	Yes	Yes	Yes	24.0
Peru	Peru-Ban	Banerjee et al. (2015)	–	2011	942	79	Yes	Yes	Yes	24.0
Philippines	Philippines-Bea1	Beam et al. (2025)	T1 Graduation Package	2018	849	250	Yes	Yes	Yes	23.0
Philippines	Philippines-Bea2	Beam et al. (2025)	T2 Graduation Package	2018	849	250	Yes	Yes	Yes	23.0
Philippines	Philippines-IPA	IPA (2022)	T1	2018	784	0	Yes	Yes	Yes	16.0
Senegal	Senegal-Bos2	Bossuroy et al. (2026)	Graduation package	2018	666	0	Yes	Yes	Yes	–

Data Appendix Table 2: Program Components by Treatment Arm *(continued)*

Country	Figure Code	Study	Treatment Arm	Implementation Year	Transfer	Cons. Support	Savings	Coaching	Training	Duration (months)
				(1)	(2)	(3)	(4)	(5)	(6)	
Senegal	Senegal-Bos3	Bossuroy et al. (2026)	Psychosocial package	2018	0	0	Yes	Yes	Yes	–
Senegal	Senegal-Bos1	Bossuroy et al. (2026)	Full package	2018	666	0	Yes	Yes	Yes	–
Somalia	Somalia-Lei	Leight et al. (2025)	Treatment	2022	–	715	Yes	Yes	Yes	–
South Sudan	South Sudan-Cho	Chowdhury et al. (2017)	TUP program	2013	344	140	Yes	Yes	Yes	24.0
Uganda	Uganda-Bea1	Beam et al. (2025)	Group coaching	2019	981	997	Yes	Yes	Yes	29.0
Uganda	Uganda-Bea2	Beam et al. (2025)	Individual coaching	2019	981	997	Yes	Yes	Yes	29.0
Uganda	Uganda-Bla1	WINGS Uganda	Group Dynamics Training	2009	440	0	Yes	Yes	Yes	6.0
Uganda	Uganda-Bla2	WINGS Uganda	No Group Dynamics Training	2009	440	0	No	Yes	Yes	6.0
Uganda	Uganda-Gui	Guilhin et al. (2022)	T3	2019	0	306	Yes	Yes	Yes	29.0
Uganda	Uganda-Mah	Mahmud et al. (2026)	Treatment	2020	111	0	Yes	Yes	Yes	24.0
Uganda	Uganda-Sed	Sedlmayr et al. (2020)	Microenterprise program	2014	82	0	Yes	Yes	Yes	9.0
Uganda	Uganda-Wan1	Wang et al. (2021)	Bridges	2012	0	0	Yes	Yes	Yes	24.0
Uganda	Uganda-Wan2	Wang et al. (2021)	Bridges PLUS	2012	0	0	Yes	Yes	Yes	24.0
Yemen, Rep.	Yemen-Bru	Brune et al. (2022)	–	2010	1,367	175	Yes	Yes	Yes	24.0
Zambia	Zambia-Bot2	Botea et al. (2023)	Full Package with consumption support	2018	286	55	Yes	Yes	Yes	9.0
Zambia	Zambia-Bot3	Botea et al. (2023)	Full Package without consumption support	2018	286	0	Yes	Yes	Yes	9.0

Notes: Transfer and consumption-support amounts are totals per treated household in 2020 PPP USD. Duration pools coaching and training because the source data do not always distinguish them cleanly; it is reported in months. "–" denotes no separately named treatment arm or genuinely unavailable duration information, as applicable. Figure codes use country plus the first three letters of the first author or organization (for example, Philippines-IPA); numbered suffixes distinguish treatment arms that would otherwise share a code. "–" in the Figure Code column denotes an arm that does not contribute a point to the manuscript figures.

Data Appendix Table 3: Observed and Predicted Program Costs by Treatment Arm

Study	Arm	Observed Cost (constant 2020 nominal USD)	Predicted Costs		Fixed Costs Included
			Predicted: Cost Covariates	Predicted: Components	
		(1)	(2)	(3)	(4)
Somalia, Leight et al. (2025)	Treatment	$3,090	–	–	Yes
Peru, Banerjee et al. (2015)	Treatment	$2,973	$915	$637	Yes
Uganda, Beam et al. (2025)	Individual coaching	$2,692	$1,401	$1,318	Yes
Uganda, Guilhin et al. (2022)	T3	$2,527	$837	$370	Yes
Uganda, Beam et al. (2025)	Group coaching	$2,401	$1,401	$1,318	Yes
Ghana, Banerjee et al. (2025)	Graduating from Ultra Poverty [GUP] with savings	$2,277	$771	$625	Yes
Ghana, Banerjee et al. (2025)	Graduating from Ultra Poverty [GUP] without savings	$2,277	$877	$922	Yes
Afghanistan, Bedoya et al. (2019)	Targeting the Ultra Poor	$1,757	$2,449	$1,428	Yes
Honduras, Banerjee et al. (2015)	Treatment	$1,549	$770	$514	Yes
Pakistan, Banerjee et al. (2015)	Treatment	$1,279	$1,047	$745	Yes
Ethiopia, Barker et al. (2024)	Treatment	$1,161	$1,152	$819	Yes
Uganda, Blattman et al. (2016)	Group Dynamics Training	$1,051	$653	$356	Yes
Uganda, Blattman et al. (2016)	No Group Dynamics Training	$951	$743	$525	Yes
Bangladesh, Beam et al. (2025)	Individual coaching	$854	$672	$1,166	Yes
Bangladesh, Beam et al. (2025)	Group coaching	$693	$672	$1,166	Yes
Nepal, Janzen et al. (2021)	Full treatment	$656	$540	$391	Yes
Mauritania, Bossuroy et al. (2026)	Full	$632	–	–	–
Nicaragua, Macours et al. (2022)	CCT and Grant	$624	$652	$885	Yes
Bangladesh, Beam et al. (2025)	Light individual coaching	$601	$672	$1,166	Yes
Philippines, Beam et al. (2025)	Individual asset, individual coaching	$576	$821	$676	Yes
Kenya, Brudevold-Newman et al. (2024)	Franchise treatment	$555	$696	$688	Yes
Mauritania, Bossuroy et al. (2026)	Capital	$546	–	–	–
Senegal, Bossuroy et al. (2026)	Full package	$456	–	–	–

Data Appendix Table 3: Observed and Predicted Program Costs by Treatment Arm *(continued)*

Study	Arm	Observed Cost (constant 2020 nominal USD)	Predicted Costs		
			Predicted: Cost Covariates	Predicted: Components	Fixed Costs Included
		(1)	(2)	(3)	(4)
Philippines, IPA (2022)	T1 [Group asset and group coaching]	$452	$318	$532	Yes
Nepal, Janzen et al. (2021)	No values based training	$438	$538	$391	Yes
Philippines, Beam et al. (2025)	Individual asset, group coaching	$422	$821	$676	Yes
Zambia, Botea et al. (2023)	Full Package with consumption support	$418	$518	$380	Yes
Senegal, Bossuroy et al. (2026)	Graduation package	$415	–	–	–
Zambia, Botea et al. (2023)	Full Package [pooled with and without CS]	$401	$750	$380	Yes
Mauritania, Bossuroy et al. (2026)	Psychosocial	$387	–	–	–
Zambia, Botea et al. (2023)	Full Package without consumption support	$384	$369	$317	Yes
Congo, Dem. Rep., Angelucci et al. (2023)	Stronger Nations Stronger Women	$376	$356	$499	Yes
Côte d'Ivoire, Marguerie et al. (2023)	Cash grant	$336	$549	$555	Yes
Kenya, Gobin et al. (2017)	REAP Full Package	$331	$290	$341	No
Burkina Faso, Bossuroy et al. (2026)	Full	$307	–	–	–
Côte d'Ivoire, Marguerie et al. (2023)	Cash grant + repayment	$297	$460	$457	Yes
Kenya, Gobin et al. (2017)	REAP Small Package	$276	$290	$341	No
Burkina Faso, Bossuroy et al. (2026)	Graduation	$268	–	–	–
Niger, Bossuroy et al. (2022)	Full arm	$260	$253	$517	Yes
Uganda, Sedlmayr et al. (2020)	Microenterprise program [A2 and B2]	$243	$121	$271	Yes
Bangladesh, Rahman et al. (2021)	Hybrid UPG [Ultra Poor Graduation]	$219	$252	$476	Unclear
Niger, Bossuroy et al. (2022)	Capital arm	$215	$253	$517	Yes
Yemen, Rep., Brune et al. (2022)	Treatment	$211	$1,286	$946	No
India, Banerjee et al. (2021)	TUP program	$208	$712	$453	Yes
Senegal, Bossuroy et al. (2026)	Psychosocial package	$190	–	–	–
Burkina Faso, Bossuroy et al. (2026)	Psychosocial	$148	–	–	–

Data Appendix Table 3: Observed and Predicted Program Costs by Treatment Arm *(continued)*

Study	Arm	Observed Cost (constant 2020 nominal USD)	Predicted Costs		Fixed Costs Included
			Predicted: Cost Covariates	Predicted: Components	
		(1)	(2)	(3)	(4)
Niger, Bossuroy et al. (2022)	Psychosocial arm	$118	$205	$397	Yes
Chad, Premand et al. (2024)	Treatment [Economic inclusion program]	$113	$165	$308	Yes
Uganda, Mahmud et al. (2026)	Treatment	$102	–	–	Yes
Afghanistan, Gibbs et al. (2020)	Empowerment program	–	–	$597	–
Bangladesh, Ahmed et al. (2026)	BRAC TUP 2007	–	–	–	–
Bangladesh, Roy et al. (2015)	Treatment [CFPR STUP]	–	–	$661	–
Burkina Faso, Bouguen et al. (2026)	T2	–	–	–	–
Burkina Faso, Bouguen et al. (2026)	T3	–	–	–	–
Egypt, Arab Rep. Assaad et al. (2026)	Full-Transfer Unrestricted + Full-Transfer Restricted	–	–	–	–
Egypt, Arab Rep., Assaad et al. (2026)	Half-Transfer	–	–	–	–
Ethiopia, Hirvonen et al. (2025)	T2 (PSNP + SPIR II graduation programming + Nurturing Care Groups)	–	–	–	–
Ethiopia, Hirvonen et al. (2025)	T2 (PSNP + SPIR II graduation programming + Nurturing Care Groups) + T3 (PSNP + SPIR II graduation programming + NCGs + maternal cash transfers)	–	–	–	–
Ethiopia, Hirvonen et al. (2025)	T3 (PSNP + SPIR II graduation programming + NCGs + maternal cash transfers)	–	–	–	–
India, Bauchet et al. (2015)	Treatment [SKS Ultra Poor Program]	–	$483	$358	Yes
Kenya, Gallardo et al. (2021)	Large Grant Village Enterprise Program	–	–	$368	–
Kenya, Gallardo et al. (2021)	Small Grant Village Enterprise Program	–	–	$285	–
Kenya, Gallardo et al. (2021)	Village Enterprise Program	–	–	$285	–
Malawi, Burchi et al. (2018)	Lump sum plus training	–	–	$400	–

Data Appendix Table 3: Observed and Predicted Program Costs by Treatment Arm *(continued)*

Study	Arm	Observed Cost (constant 2020 nominal USD)	Predicted Costs		Fixed Costs Included
			Predicted: Cost Covariates	Predicted: Components	
		(1)	(2)	(3)	(4)
Malawi, Burchi et al. (2018)	Training only	–	–	$376	–
Malawi, Poll et al. (2025)	Treatment	–	–	–	–
South Sudan, Chowdhury et al. (2017)	TUP program	–	–	$423	–
Uganda, Wang et al. (2021)	Bridges	–	–	$255	–
Uganda, Wang et al. (2021)	Bridges PLUS	–	–	$255	–

Notes: All three cost columns are expressed in constant 2020 nominal USD per household. The observed-cost column uses reported nominal costs where available and reviewed nominal equivalents for PPP-only source costs. The two prediction columns report separate full-model (cost metadata and components) and component-only predictions; missing entries indicate unavailable costs or predictions and are not replaced by values from another column. These nominal display measures are distinct from the PPP costs used in treatment-effect comparisons and benefit-cost calculations. The Zambia pooled full-package contrast is retained here for traceability but excluded from unique-arm descriptive cost summaries when both constituent full-package costs are available.

Data Appendix Table 4: Cost-Benefit Analysis: All Studies

Sorted by Benefit/Cost Ratio under estimated persistence

Study / Arm	Linear decay (1)	Estimated persistence (2)	Permanent persistence (3)	Nominal cost (4)
Uganda, Mahmud - Treatment	20.75 [15.55, 25.97]	18.30 [10.16, 26.38]	84.60 [61.12, 108.48]	$102
Chad, Premand	5.82 [1.99, 9.63]	14.60 [5.69, 23.34]	25.25 [8.63, 41.75]	$113
Bangladesh, Rahman	5.00 [2.76, 7.31]	12.48 [7.05, 18.00]	26.01 [14.35, 38.04]	$219
Niger, Bossuroy - Psychosocial Package	6.47 [3.12, 9.85]	11.70 [2.40, 21.30]	28.07 [13.52, 42.72]	$118
Niger, Bossuroy - Full Package	3.98 [2.45, 5.55]	9.31 [4.80, 13.81]	17.27 [10.63, 24.08]	$260
Senegal, Bossuroy - Psychosocial Package	6.16 [1.60, 10.71]	9.19 [1.81, 16.76]	16.79 [0.77, 32.47]	$190
Niger, Bossuroy - Capital Package	2.40 [0.96, 3.86]	8.23 [3.16, 13.28]	10.42 [4.16, 16.72]	$215
Philippines, DOLE Graduation - Individual Asset, Group Coaching	3.98 [1.76, 6.23]	6.41 [3.10, 9.78]	14.89 [6.60, 23.33]	$422
Philippines, IPA - Coach Group + Asset Group	3.88 [1.02, 6.79]	6.33 [2.09, 10.64]	14.53 [3.83, 25.40]	$452
Uganda, WINGS - No Group Dynamics Training	2.63 [1.53, 3.80]	5.87 [3.56, 8.36]	13.69 [7.95, 19.77]	$951
India, Banerjee (2021) - TUP Program	7.34 [5.34, 9.31]	5.86 [2.85, 8.82]	16.68 [9.57, 23.74]	$208
Senegal, Bossuroy - Full Package	2.86 [0.82, 4.91]	5.29 [1.97, 8.62]	5.72 [-1.56, 13.03]	$456
Uganda, Sedlmayr - Microenterprise Program	1.11 [0.26, 1.95]	4.59 [1.01, 8.19]	4.25 [-0.11, 8.65]	$243
Kenya, Gobin - REAP Full Package	-0.68 [-1.59, 0.21]	4.49 [0.50, 8.65]	-3.54 [-8.30, 1.12]	$331
Philippines, DOLE Graduation - Individual Asset, Individual Coaching	2.70 [0.88, 4.55]	4.41 [1.68, 7.15]	10.12 [3.30, 17.01]	$576
Yemen, Brune	2.90 [-3.08, 8.79]	4.30 [0.00, 11.83]	9.59 [-10.20, 29.09]	$211
Senegal, Bossuroy - Graduation Package	2.98 [0.96, 5.02]	4.29 [1.16, 7.46]	8.31 [0.92, 15.90]	$415
Kenya, Gobin - REAP Small Package	0.54 [-0.59, 1.65]	3.64 [0.11, 7.58]	2.79 [-3.08, 8.58]	$276
Zambia, Botea - Full Package Without Consumption Support	3.08 [1.94, 4.21]	3.12 [1.63, 4.63]	10.52 [6.00, 15.13]	$384
Zambia, Botea - Full Package With Consumption Support	2.12 [1.02, 3.22]	2.89 [1.40, 4.36]	6.59 [2.19, 10.94]	$418
DRC, Angelucci	0.42 [0.23, 0.62]	2.66 [0.47, 4.87]	1.74 [0.77, 2.72]	$376

Continued on next page

Study / Arm	Linear decay (1)	Estimated persistence (2)	Permanent persistence (3)	Nominal cost (4)
Mauritania, Bossuroy - Psychosocial	1.29 [-1.03, 3.65]	2.55 [0.00, 6.07]	4.84 [-3.85, 13.66]	$387
Bangladesh, Beam - Individual Coaching	2.46 [1.25, 3.67]	2.44 [1.16, 3.73]	6.70 [3.40, 10.00]	$854
Bangladesh, Beam - Group Coaching	2.40 [1.12, 3.71]	2.35 [0.98, 3.75]	6.54 [3.05, 10.11]	$693
Mauritania, Bossuroy - Full	1.15 [-0.16, 2.45]	2.09 [0.27, 4.04]	4.30 [-0.59, 9.19]	$632
Somalia, Leight - Treatment	1.14 [0.96, 1.31]	2.02 [1.69, 2.36]	4.92 [4.16, 5.68]	$3,090
Uganda, Graduating to Resilience - Group Coaching	1.08 [0.87, 1.28]	1.95 [1.56, 2.34]	4.67 [3.77, 5.57]	$2,401
Bangladesh, Beam - Light Individual Coaching	2.03 [0.51, 3.58]	1.93 [0.45, 3.57]	5.53 [1.38, 9.75]	$601
Burkina Faso, Bossuroy - Graduation	0.62 [-2.10, 3.29]	1.92 [0.00, 5.89]	2.31 [-7.88, 12.32]	$268
Burkina Faso, Bossuroy - Psychosocial	-0.29 [-4.79, 4.18]	1.89 [0.00, 8.01]	-1.09 [-17.93, 15.64]	$148
Nicaragua, Macours - Cct And Grant	1.08 [0.50, 1.66]	1.89 [0.96, 2.81]	4.05 [1.86, 6.22]	$624
Mauritania, Bossuroy - Capital	0.89 [-1.05, 2.81]	1.76 [0.00, 4.61]	3.31 [-3.92, 10.50]	$546
Afghanistan, Bedoya	0.93 [0.59, 1.29]	1.74 [1.11, 2.37]	4.04 [2.55, 5.58]	$1,757
Uganda, Graduating to Resilience - Individual Coaching	0.90 [0.74, 1.05]	1.63 [1.32, 1.94]	3.89 [3.19, 4.58]	$2,692
Kenya, Brudevold-Newman - Franchise Package	0.43 [-0.66, 1.49]	1.62 [0.00, 3.67]	1.87 [-2.84, 6.48]	$555
Ethiopia, Banerjee/Barker	0.94 [0.58, 1.29]	1.27 [0.63, 1.89]	2.15 [0.83, 3.47]	$1,161
Ghana, Banerjee - GUP Without Savings	0.78 [0.27, 1.31]	1.14 [0.02, 2.41]	3.21 [0.99, 5.49]	$2,277
Pakistan, Banerjee	0.52 [0.19, 0.86]	1.09 [0.43, 1.78]	1.81 [0.23, 3.39]	$1,279
Ghana, Banerjee - GUP With Savings	0.50 [0.09, 0.91]	0.79 [0.00, 2.01]	1.87 [-0.09, 3.82]	$2,277
Honduras, Banerjee	-0.12 [-0.67, 0.43]	0.71 [0.00, 1.81]	-0.92 [-3.53, 1.65]	$1,549
Peru, Banerjee	0.39 [0.08, 0.71]	0.68 [0.04, 1.40]	1.56 [0.13, 3.00]	$2,973
Burkina Faso, Bossuroy - Full	-0.55 [-2.84, 1.67]	0.56 [0.00, 3.39]	-2.05 [-10.62, 6.27]	$307
Uganda, Guilhin - T3	0.13 [0.09, 0.18]	0.36 [0.17, 0.55]	0.58 [0.38, 0.78]	$2,527
Egypt, Arab Rep., Assaad: 1/2		No cost data		
Egypt, Arab Rep., Assaad - Half-Transfer		No cost data		

Continued on next page

Data Appendix

Study / Arm	Linear decay	Estimated persistence	Permanent persistence	Nominal cost
	(1)	(2)	(3)	(4)
India, Bauchet		No cost data		
Malawi, Burchi - Lump Sum Plus Training		No cost data		
Malawi, Burchi - Training Only		No cost data		
South Sudan, Chowdhury - Tup		No cost data		
Ethiopia, Hirvonen - T2 (Psnp + Spir Ii Graduation Programming + Nurturing Care Groups)		No cost data		
Ethiopia, Hirvonen - T3 (Psnp + Spir Ii Graduation Programming + Ncgs + Maternal Cash Transfers)		No cost data		
Bangladesh, Bandiera/Ahmed		No CBA (unknown)		

Notes: Benefit-cost ratios use consumption benefits only and reported per-household program costs, discounted at 5 percent per year. Cells report posterior medians with 95 percent credible intervals in brackets. Column (1) applies linear decay of 12.5 percent per year, corresponding to 50 percent decay after four years and full decay after eight. Column (2) projects benefits over a common 10-year horizon integrating over the full posterior distribution of the slope from the pooled Graduation consumption follow-up model and floors projected consumption effects at zero. Column (3) holds the final observed annual consumption effect constant forever, beginning one year after the final observed follow-up. The final column displays costs in constant 2020 nominal USD, using reported nominal costs where available and PPP-backconverted nominal equivalents otherwise; benefit-cost ratios continue to use reported costs in constant 2020 PPP USD as their analytical denominator. Arms without reported PPP costs are listed last and excluded from CBA averages.

Appendix: Papers Included in the Meta-Analysis

The 49 study records in Data Appendix Table 1 map to the 41 distinct source papers listed below. Multi-country papers and papers used for more than one study record appear once. This list is restricted to papers that contribute treatment-effect estimates to the meta-analysis; background and methodological references are listed separately in the main References section.

Ahmed, S., Burgess, R., Chowdhury, S., Haque, S. T., Hossain, M. A. N., Matin, I., and Sulaiman, M. Initial evidence on the intergenerational effects of the ultra-poor graduation program in Bangladesh. G^2LM—LIC Working Paper 105, July 2026.

Angelucci, M., Heath, R., and Noble, E. Multifaceted programs targeting women in fragile settings: Evidence from the Democratic Republic of Congo. *Journal of Development Economics*, 164:103146, 2023. doi: 10.1016/j.jdeveco.2023.103146. Paper link. DRC — DRC-Ang.

Assaad, R., Osman, A., Parienté, W., and Valente, C. How big does a "big push" need to be? Evidence from randomizing asset transfer size. Working paper, April 24, 2026.

Bandiera, O., Burgess, R., Das, N., Gulesci, S., Rasul, I., and Sulaiman, M. Labor markets and poverty in village economies. *The Quarterly Journal of Economics*, 132(2):811–870, 2017. doi: 10.1093/qje/qjx003. Paper link. Bangladesh — Bangladesh-Ban.

Banerjee, A., Duflo, E., Chattopadhyay, R., and Shapiro, J. Targeting the hard-core poor: An impact assessment. Technical report, Poverty Action Lab / J-PAL, 2011. Working Paper. Paper link. India — India-Ban.

Banerjee, A., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Parienté, W., Shapiro, J., Thuysbaert, B., and Udry, C. A multifaceted program causes lasting progress for the very poor: Evidence from six countries. *Science*, 348(6236):1260799, 2015. doi: 10.1126/science.1260799. Paper link. Ethiopia — Ethiopia-Ban. Ghana — Ghana-Ban1–2. Honduras — Honduras-Ban. India — India-Ban. Pakistan — Pakistan-Ban. Peru — Peru-Ban.

Banerjee, A., Duflo, E., and Sharma, G. Long-term effects of the targeting the ultra poor program. *American Economic Review: Insights*, 3(4):471–486, 2021a. doi: 10.1257/aeri.20200667. Originally circulated as NBER Working Paper No. 28074, November 2020. Paper link. India — India-Ban.

Banerjee, A., Karlan, D., Osei, R. D., Trachtman, H., and Udry, C. Unpacking a multi-faceted program to build sustainable income for the very poor. *Journal of Development Economics*, 155:102781, 2021b. doi: 10.1016/j.jdeveco.2021.102781. Originally circulated as NBER Working Paper No. 24271, 2018. Paper link. Ghana — Ghana-Ban1–2.

Barker, N., Karlan, D., Udry, C., and Wright, K. The fading treatment effects of a multifaceted asset-transfer program in ethiopia. *American Economic Review: Insights*, 6(2):277–294, 2024. doi: 10.1257/aeri.20230156. Paper link. Ethiopia — Ethiopia-Ban.

Bauchet, J., Morduch, J., and Ravi, S. Failure vs. displacement: Why an innovative anti-poverty program showed no net impact in South India. *Journal of Development Economics*, 116:1–16, 2015. doi: 10.1016/j.jdeveco.2015.03.005. Paper link. India — India-Bau.

Beam, E. A., Brune, L., Das, N., Dercon, S., Goldberg, N., Haque, R., Karlan, D., Khan, M., Parkerson, D., Pople, A., Sawada, Y., Udry, C., and Zizzamia, R. Group versus individual coaching for rural social protection programs: Evidence from uganda, philippines, and bangladesh. Working Paper 34309, National Bureau of Economic Research, Oct. 2025. Paper link. Bangladesh — Bangladesh-Bea1–3. Philippines — Philippines-Bea1–2. Uganda — Uganda-Bea1–2.

Bedoya, G., Coville, A., Haushofer, J., Isaqzadeh, M., and Shapiro, J. P. No household left behind: Afghanistan targeting the ultra poor impact evaluation. Working Paper 25981, NBER, 2019. Paper link. Afghanistan — Afghanistan-Bed.

Blattman, C., Green, E. P., Jamison, J., Lehmann, M. C., and Annan, J. The returns to microenterprise support among the ultrapoor: A field experiment in postwar Uganda. *American Economic Journal: Applied Economics*, 8(2):35–64, 2016. doi: 10.1257/app.20150023. Paper link. Uganda — Uganda-Bla1–2.

Bossuroy, T., Goldstein, M., Karimou, B., Karlan, D., Kazianga, H., Parienté, W., Premand, P., Thomas, C., Udry, C., Vaillant, J., and Wright, K. Tackling psychosocial and capital constraints to alleviate poverty. *Nature*, 605:291–297, 2022. doi: 10.1038/s41586-022-04647-8. Originally circulated as World Bank Policy Research Working Paper No. 9562 (2021) under the title "Pathways out of Extreme Poverty: Tackling

Psychosocial and Capital Constraints with a Multi-faceted Social Protection Program in Niger". Paper link. Niger — Niger-Bos1–3.

Bossuroy, T., Karlan, D., Kazianga, H., Parienté, W., Premand, P., Udry, C., Vaillant, J., and Wright, K. Multi-Faceted Economic Inclusion in Burkina Faso.

Bossuroy, T., Karlan, D., Kazianga, H., Parienté, W., Premand, P., Udry, C., Vaillant, J., and Wright, K. Multi-Faceted Economic Inclusion in Mauritania.

Bossuroy, T., Karlan, D., Kazianga, H., Parienté, W., Premand, P., Udry, C., Vaillant, J., and Wright, K. Multi-Faceted Economic Inclusion in Senegal.

Botea, I., Brudevold-Newman, A., Goldstein, M., Low, C., and Roberts, G. Supporting women's livelihoods at scale: Evidence from a nationwide multi-faceted program. Working Paper 31625, National Bureau of Economic Research, 2023. Paper link. Zambia — Zambia-Bot1–3.

Bouguen, A. and Dillon, A. Beyond poverty reduction: Evidence from a multifaceted program on poverty, nutrition and child development. *Journal of Development Economics*, 178:103596, 2026. doi: 10.1016/j.jdeveco.2025.103596. Paper link.

Brudevold-Newman, A., Honorati, M., Ipapa, G., Jakiela, P., and Ozier, O. A firm of one's own: Experimental evidence on credit constraints and occupational choice. *The Review of Economics and Statistics*, 2024. doi: 10.1162/rest_a_01529. Paper link. Kenya — Kenya-Bru.

Brune, L., Karlan, D., Kurdi, S., and Udry, C. Social protection amidst social upheaval: Examining the impact of a multi-faceted program for ultra-poor households in yemen. *Journal of Development Economics*, 155: 102780, 2022. ISSN 0304-3878. doi: 10.1016/j.jdeveco.2021.102780. Paper link. Yemen — Yemen-Bru.

Burchi, F. and Strupat, C. Unbundling the impacts of economic empowerment programmes: Evidence from Malawi. Discussion Paper 32/2018, German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE), Bonn, 2018. Paper link. Malawi — Malawi-Bur1–2.

Chowdhury, R., Collins, E., Ligon, E., and Munshi, K. Valuing assets provided to low-income households in South Sudan, Dec. 2016. URL Working paper dated December 15, 2016. Paper link. South Sudan — South Sudan-Cho.

Gibbs, A., Corboz, J., Chirwa, E., Mann, C., Karim, F., Shafiq, M., Mecagni, A., Maxwell-Jones, C., Noble, E., and Jewkes, R. The impacts of combined social and economic empowerment training on intimate partner violence, depression, gender norms and livelihoods among women: an individually randomised controlled trial and qualitative study in afghanistan. *BMJ Global Health*, 5:e001946, 2020. doi: 10.1136/bmjgh-2019-001946. Paper link. Afghanistan — Afghanistan-Gib.

Gobin, V. J., Santos, P., and Toth, R. Poverty graduation with cash transfers: A randomized evaluation. Monash Economics Working Paper 23-16, Monash University, Department of Economics, 2016. Paper link. Kenya — Kenya-Gob1–2.

Hirvonen, K., Leight, J., Gilligan, D. O., Mesfin, H. M., Mulford, M., and Tesfaye, H. The impact of a nutrition-sensitive graduation model program on child nutrition: Experimental evidence from Ethiopia. IFPRI Discussion Paper 02391, International Food Policy Research Institute, December 2025. Paper link.

Innovations for Poverty Action Philippines. Impact evaluation of the graduation of the ultra-poor pilot in the Philippines: Report — endline survey. Technical report, Innovations for Poverty Action Philippines, Manila, Apr. 2022. Paper link. Philippines — Philippines-IPA1–3.

Janzen, S. A., Magnan, N., Sharma, S., and Thompson, W. M. Pay it forward: Impacts of a rural livelihoods program with built-in spillovers. Technical report, University of Illinois and University of Georgia, June 2021. URL Working paper dated June 14, 2021. Paper link. Nepal — Nepal-Jan1–2.

Jimenez Gallardo, M. A., Kananu Ntonjira, W., Lazicky, C., McManus, J., Njogu-Ndongwe, F., and Pignon, C. Village enterprise development impact bond evaluation findings. Technical report, IDinsight, Mar. 2022. URL Paper link. Kenya — Kenya-Gal1–3.

Leight, J., Hirvonen, K., Karachiwalla, N., and Rakshit, D. Displacement and development: Evidence from a graduation program for Somalia's ultra-poor. IFPRI Discussion Paper 02356, International Food Policy Research Institute, September 2025. Paper link.

Macours, K., Premand, P., and Vakis, R. Transfers, diversification and household risk strategies: Can productive safety nets help households manage climatic variability? *The Economic Journal*, 2022. doi: 10.1093/ej/ueac018. Paper link. Nicaragua — Nicaragua-Mac1.

Mahmud, M. and Riley, E. "We're all in this together": Experimental evidence from a universal livelihoods

program. Working paper, April 6, 2026.

Malaeb, B. and Uzor, E. Multidimensional impact evaluation: A randomized control trial on conflict-affected women in Northern Uganda. Technical report, Centre for the Study of the Economies of Africa, Abuja, 2023. Paper link. Uganda — Uganda-Bla1–2.

Marguerie, A. and Premand, P. Savings facilitation or capital injection? Impacts and spillovers of livelihood interventions in post-conflict Côte d'Ivoire. Policy Research Working Paper 10563, World Bank, Washington, DC, Sept. 2023. Paper link. Côte d'Ivoire — Côte d'Ivoire-Mar1–2.

Poll, M. Poverty alleviation at high frequency. Job market paper, Brown University, November 26, 2025. Paper link.

Premand, P. and Schnitzer, P. Impacts and spillovers of a low-cost multi-faceted economic inclusion program in Chad. Policy Research Working Paper 11084, World Bank, 2025. https://documents.worldbank.org/curated/en/099743503122576854/pdf/IDU-0d27fcb3-d43a-466b-ad3e-6f1bdcac0647.pdf. Paper link. Chad — Chad-Pre.

Rahman, A., Bhattacharjee, A., and Das, N. A good mix against ultra-poverty? evidence from a randomized controlled trial (RCT) in Bangladesh. *Review of Development Economics*, 2021. doi: 10.1111/rode.12809. Source file provided as local filename "270 Rehman et al A good mix against ultra-poverty.pdf"; online link not available in spreadsheet. Paper link. Bangladesh — Bangladesh-Rah.

Roy, S., Ara, J., Das, N., and Quisumbing, A. R. Flypaper effects in transfers targeted to women: Evidence from BRAC's targeting the ultra poor program in Bangladesh. *Journal of Development Economics*, 117: 1–19, 2015. doi: 10.1016/j.jdeveco.2015.06.004. Paper link. Bangladesh — Bangladesh-Roy.

Sedlmayr, R., Shah, A., and Sulaiman, M. Cash-plus: Poverty impacts of alternative transfer-based approaches. *Journal of Development Economics*, 144:102418, 2020. doi: 10.1016/j.jdeveco.2019.102418. Originally circulated as CSAE Working Paper WPS/2017-15 (2018) under the title "Cash-Plus: Poverty Impacts of Transfer-Based Intervention Alternatives". Paper link. Uganda — Uganda-Sed.

Technical and Operational Performance Support (TOPS) Uganda Graduation Randomized Control Trial Associate Award. Endline report of the resilience food security activity graduating to resilience in uganda, cohort 1. Technical report, The TOPS Program, Washington, DC, Oct. 2022. Paper link. Uganda — Uganda-Gui.

Wang, J. S.-H., Malaeb, B., Ssewamala, F. M., Neilands, T. B., and Brooks-Gunn, J. A multifaceted intervention with savings incentives to reduce multidimensional child poverty: Evidence from the Bridges study (2012–2018) in rural Uganda. *Social Indicators Research*, 158(3):947–990, 2021. doi: 10.1007/s11205-021-02712-9. Paper link. Uganda — Uganda-Wan1–2.