100Weeks evidence review

Report prepared for HereWeGrow

June 2025

laterite
DATA | RESEARCH | ANALYTICS

ACKNOWLEDGEMENTS

We would like to thank the teams at **100Weeks** and **HereWeGrow** for their collaboration and openness throughout the course of this project. Their deep knowledge of the program and commitment to improving the lives of participants have been instrumental to this analysis.

We acknowledge the participants of the 100Weeks program whose experiences and data form the foundation of this report. Their voices continue to guide efforts to design more impactful and inclusive development programs.

AI USAGE DISCLAIMER

At Laterite, we may use Artificial Intelligence (AI) tools to support and enhance our research processes, including tasks such as data transcription, analysis, and report drafting. All AI-generated outputs are reviewed, verified, and refined by qualified team members with expertise in relevant subject matter and local context to ensure quality and accuracy.

We prioritize transparency and adhere to strict privacy safeguards, ensuring that no personally identifiable information (PII) is processed by AI systems. We use only pre-approved AI tools that do not retain or train on our data. AI is never used as a substitute for expert judgment, and all final deliverables reflect the expertise and oversight of our team.

For more information on our AI usage practices, please refer to our internal AI policy. In addition to this general disclaimer, the use of AI in reviewing documents is explained in “2.4 External Research Documents”.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Executive Summary

100Weeks is a 100-week program that provides weekly unconditional cash transfers, financial training, and group savings support to help women escape poverty and build lasting financial independence. Women receive €8 per week via mobile money, enabling them to meet basic needs and invest in small businesses. Weekly coaching and participation in savings groups foster long-term planning, resilience, and income generation beyond the program.

This report presents the analytical work conducted by Laterite for 100Weeks and HereWeGrow. It provides a comprehensive assessment of the 100Weeks program, beginning with an overview of the program design and situating it within the broader context of existing research. The report focuses on four key areas: (1) a review of the evidence base surrounding cash transfer programs, (2) a heterogeneity analysis exploring how program effects vary across participant groups, (3) a benchmarking exercise projecting potential monetary consumption gains based on an external study for Uganda, and (4) an estimate of the program's return on investment (ROI) based on the benchmarking. Together, these components offer a holistic view of the program's effectiveness and potential impact. The main findings are summarized below.

1) Data Quality and MEL Practice

100Weeks uses a comprehensive monitoring system to track progress during and after the implementation of their standard program intervention cycle. Their approach to MEL is adaptive and learning oriented. 100Weeks treats monitoring, evaluation , and learning (MEL) as a living process. A purpose-built system tracks participants from the program's launch to four years after completion. Trained enumerators gather data at baseline, at three interim checkpoints, at endline (after 100weeks), and during follow-up surveys; using cross-nationally comparable and validated instruments such as the Multidimensional Poverty Index (MPI) and the Household Dietary Diversity Score (HDDS). These metrics, complemented by indicators on mental health and intra-household decision-making power, give 100Weeks a multidimensional view of poverty and generate the feedback it needs to continuously refine the program.

Several external studies complement the internal MEL data shaping our current understanding of the evidence base around the 100Weeks model. These studies include seven theses, a research report and a peer-reviewed journal article featuring qualitative, quantitative and combined approaches. The studies enrich the understanding of program impact. In one study, the 100Weeks program is compared to other similar interventions. In another study, the program participants are compared to a group of non-participants, established at endline.

Key gaps in the current evidence base include insufficient understanding of how participants make investment decisions and how training shapes behavioral change. The key limitation of most of these studies is that they do not include a valid counterfactual, which hampers the ability to attribute observed changes in poverty reduction directly to the program.

2) Pre-Post MPI Changes vs. external research

Internal monitoring data shows large reductions in multidimensional poverty of program participants over time:

- MPI scores dropped (indicating lower levels of poverty) from 0.38 to 0.21 on average (-0.17).
- Poverty rates dropped from 69% to 17% using the standard cutoff poverty cut-off (>0.33).
- In Uganda, poverty fell from 57% to 8% for program participants.

However, these large differences are likely an overestimation of the true program effects, as this is not based on a comparison with a counterfactual group. The externally commissioned research that employs more rigorous designs (e.g., RCTs and matched comparisons) therefore reports much smaller changes in MPI, approximately -0.04 points. However, these studies also do not inform us about the true program effect, due to limitations stemming from the characteristics of the samples used and the study design. In conclusion, the big pre-post differences in MPI might be an overestimation of the true program effect as poverty declines across the board, but the smaller estimation of the MPI in external research documents is likely to be an underestimation of the true program effects.

Even though external studies offer limited insights on the causal net effect of the program on its relevant outcome indicators, they provide a useful contribution to understanding the Theory of Change (ToC) and key mechanisms such as increased savings, income diversification, and business/farm investments.

3) Heterogeneity in MPI Reductions

The heterogeneity analysis of MPI reductions highlights that certain groups experience greater improvements than others. Households that were poorer at baseline saw the largest reductions in multidimensional poverty, confirming that initial poverty status is a strong predictor of impact. Country-level differences were also significant, with Uganda and Rwanda showing the most substantial declines in MPI scores. When comparing livelihood types, subsistence and conservation farmers experienced greater poverty reductions than cocoa and coffee farmers, even after adjusting for other covariates. Additional variation was observed by household composition and education: households without a partner improved more than those with a partner, and households with a literate head experienced greater MPI improvements compared to those led by someone without formal education. These findings reinforce the importance of tailoring interventions to reach and support the most vulnerable groups.

The component-level analysis of the MPI reveals that improvements in nutrition is the primary driver of reductions in multidimensional poverty, accounting for over 50% of the observed variation in MPI decline. Other meaningful contributors include gains in asset ownership, increased access to electricity, and improved school attendance among school-aged children. In contrast, components such as cooking fuel and years of schooling contributed minimally to overall poverty reduction. The analysis also highlights country-specific differences in the relative importance of various MPI components, underscoring the need for context-sensitive program design.

4) Return on Investment

The analysis estimates a baseline Return on Investment (ROI) of 2.96 for the 100Weeks program, assuming average predicted consumption gains, a 10% discount rate, and a 3% annual decay in effects. Sensitivity testing across a range of plausible discount rates (5–15%) and decay rates (+3% to -15%) produced ROI estimates ranging from 1.91 to 4.4. In all cases, ROI remained above 1, indicating that the program's benefits outweigh its costs under a variety of assumptions.

However, the absence of a counterfactual limits the strength of causal claims, and assumptions about long-term benefits introduce further uncertainty. Despite these limitations, the analysis highlights two clear priorities: first, to improve measurement by including monetary consumption or income data alongside MPI; and second, to strengthen targeting of poorer households, who show the largest gains. These improvements would enhance both the accuracy of future evaluations and the program's cost-effectiveness.

1 Introduction

100Weeks is a 100-week program that combines unconditional weekly cash transfers, structured training, and group savings mechanisms to help women escape poverty and build long-term financial independence. Each participant receives 8 EUR per week via mobile money, allowing them to immediately address essential needs such as food, school fees, and housing. As the program progresses, participants are encouraged to invest in income-generating activities and small businesses, promoting a shift from short-term relief to sustainable livelihoods.

A key feature of the program is weekly training sessions led by local coaches, covering topics such as financial literacy, entrepreneurship, life skills, and climate-smart agriculture. These sessions aim to build decision-making and planning capacity. Participants also form Village Savings and Loan Associations (VSLAs), using part of their transfer to develop savings and access small loans. These groups support investments during the program and persist beyond its end, ensuring continued financial resilience after cash support stops.

In this report of the 100Weeks model for HereWeGrow, we present the evidence review, the heterogeneity analysis, the benchmarking to project how income effects of the 100Weeks program might materialize, and the return on investment.

Chapter 2, the Evidence Review, will critically examine the quality and results of research reports on 100Weeks and 100Weeks' current monitoring, evaluation, and learning (MEL) practices. This analysis will reveal what we currently do and do not know about the program effects, especially regarding the Multidimensional Poverty Index (MPI).

Chapter 3, Heterogeneity Analysis, will be a detailed review of the data. The heterogeneity analysis investigates which recipient profiles experience the biggest benefits from the program. By examining how outcomes vary across different subgroups (such as start year, country and baseline poverty levels), the analysis uncovers key factors that drive success for various participants. We will focus on treatment group data, baseline (BL) and endline (EL) (after 100 weeks) to assess heterogeneous impacts on key outcomes, including the MPI and food security measures.

Chapter 4, Benchmarking, will link the 100Weeks Ugandan participants dataset to a an externally collected sample of Ugandan coffee farmers. Using harmonized indicators, we apply the relationship between MPI reduction and consumption observed in Raising the Village (RTV) data to the 100Weeks BL and EL data, predicting monetary consumption. This benchmarking exercise provides input for the return on investment chapter that follows.

Chapter 5, Return on Investment, will provide rough estimates of the efficiency of the 100Weeks program by comparing the predicted monetary value of consumption gains obtained in Chapter 4 to the program's costs per treated beneficiary. The analysis projects benefits over a 10-year period, applying a conservative depreciation rate, and expresses all values in 2017 PPP USD for comparability.

2 Evidence review

2.1 Objectives

This evidence review will assess the quality and findings of research reports about 100Weeks and 100Weeks' current MEL practices to evaluate both their methodological rigor and the conclusions they provide about program effectiveness.

This chapter begins by giving an overview of the 100Weeks model (2.2) and examining the data quality of their monitoring of program participants, including a snapshot of 100Weeks' current MEL practice (2.3). Then, the observed changes in MPI during the program based on that data will be presented (2.5). It is important to acknowledge that without a control group, observed changes between the program's start and end cannot be directly attributed to the program. A contextualization of the observed MPI changes can partly mitigate that knowledge gap. Additionally, we will draw upon research reports about the 100Weeks program, which might offer a stronger basis for understanding program impact (2.4). The quality of the documents is assessed, key take-away messages are presented and contributions to the ToC from those documents are visualized. Key results emerging from the research documents will be compared to the pre-post changes we observe in the data (2.6). Finally, the key findings from these areas will be synthesized in the conclusion (2.7).

2.2 The 100Weeks model

The 100Weeks program is centered on empowering families to move out of poverty sustainably and on their own terms. This is measured through an MPI score that is less than 0.33. By providing cash grants focused on improving living standards, health outcomes, social capital, competencies, and children's school attendance, the program aims to create lasting changes that enable families to achieve financial resilience and independence. The combined effect of these efforts leads to improved well-being, better decision-making within families, enhanced social support, and the development of critical entrepreneurial and life skills, ultimately contributing to long-term poverty reduction and economic empowerment.

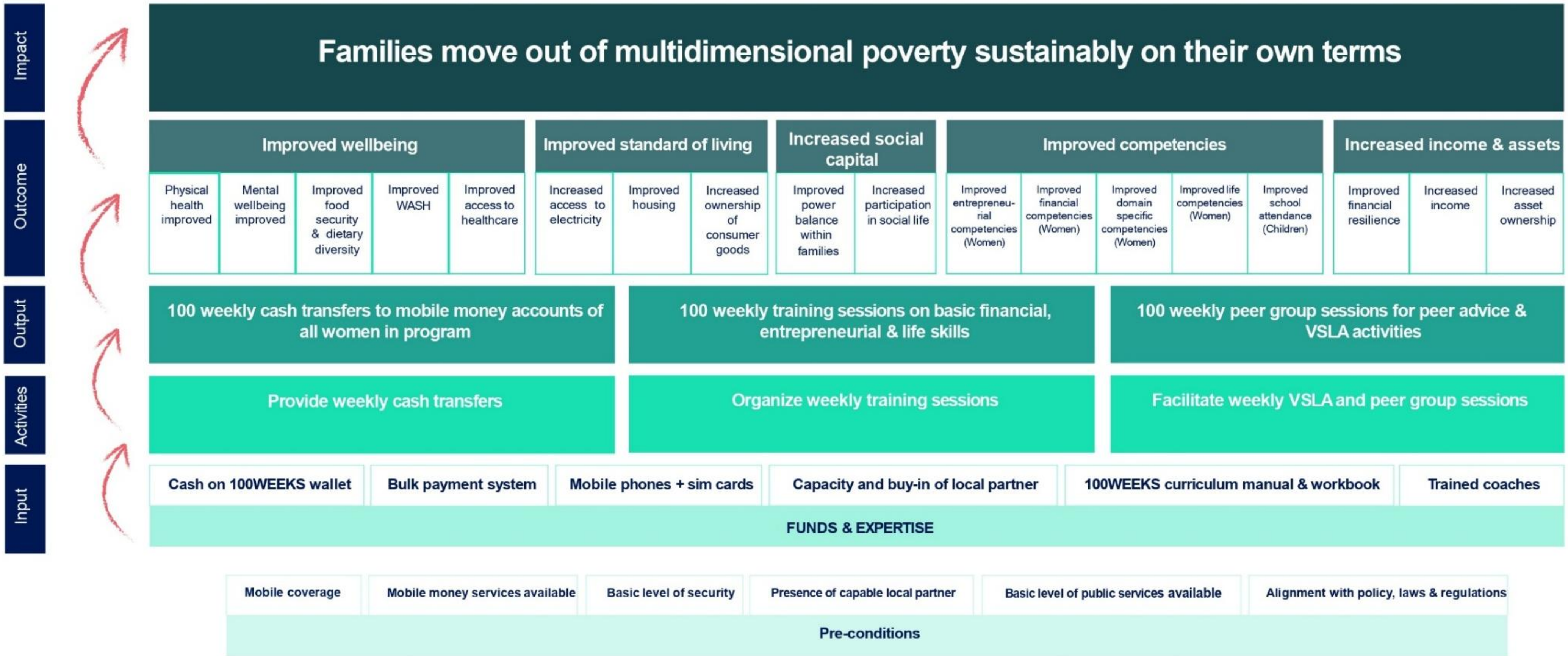
100Weeks provides unconditional cash transfers, weekly training, and group savings mechanisms for a 100-week period (see Figure 1 below). Each participant – primarily women – receives a weekly transfer of 8 EUR via mobile money, providing them with immediate financial relief without any conditions on its use. In the initial weeks, most participants use the funds to cover essential needs, such as food, school fees, and housing improvements. As they progress through the program, they are encouraged to invest in income-generating activities and small businesses, aiming for a transition toward financial independence.

Another essential component of the program is the formation of Village Savings and Loan Association (VSLA) groups, where participants set aside a portion of their weekly cash transfers. These savings groups provide access to small loans, allowing participants to make larger investments when needed. By the end of the 100-week cycle, the objective is that participants have built substantial savings, diversified their income sources, and developed financial management skills. Importantly, these savings groups continue functioning beyond the program, ensuring participants maintain financial stability and independence long after direct cash support ends.

Weekly training sessions are delivered by locally trained coaches. These sessions follow a structured curriculum covering financial literacy, entrepreneurship, personal empowerment, life skills, and climate-smart agriculture. The training aims to help participants develop economic resilience, better decision-making skills, and long-term planning abilities, ensuring they can maximize the impact of their financial resources. Below is a breakdown of the topics throughout the weekly training sessions:

1. **Introduction (Weeks 1-5):** Participants are introduced to the 100Weeks program, get to know each other, and engage in group activities to build a sense of community. They also participate in planning exercises.
2. **Personal Empowerment (Weeks 6-20):** These sessions focus on self-awareness, confidence-building, personal growth, behavior change, and developing personal empowerment plans. Participants address topics like jealousy, goal setting, and positive thinking.
3. **Financial Literacy (Weeks 21-35):** This module helps participants develop better financial habits. It covers money management, savings goals, income, and spending, distinguishing between needs and wants, budgeting, and conflict resolution related to money.
4. **Entrepreneurship (Weeks 36-55):** The focus shifts to business development, market research, generating and selecting business ideas, calculating costs and profit, marketing, record-keeping, and business action plans. Participants are equipped with the skills to start and sustain small businesses.
5. **Climate-Smart Agriculture (Weeks 56-65)**: This module teaches sustainable farming practices, including intercropping, crop rotation, climate change adaptation, soil fertility, pest control, and post-harvest management.
6. **Life Skills (Weeks 66-85)**: Participants receive training on health, nutrition, family planning, hygiene, sanitation, gender-based violence, and mental health. These sessions aim to improve their overall well-being.
7. **Flexible Sessions (Weeks 86-95):** This period allows for additional training or revisiting key topics based on participant needs.
8. **End of 100Weeks (Weeks 96-100):** The final weeks focus on long-term financial planning, budgeting, group dynamics, and setting future goals to ensure sustainability beyond the program.

Figure 1: 100Weeks' Theory of Change



Source: 100Weeks

Figure 1 provides an overview of the program and the Theory of Change (ToC), detailing the activities conducted by 100Weeks, the outputs, measured outcomes, and the hypothesized links between them. The ToC maps the causal pathway from foundational preconditions (such as mobile coverage and security), inputs to the program's activities and outputs to the intended outcomes.

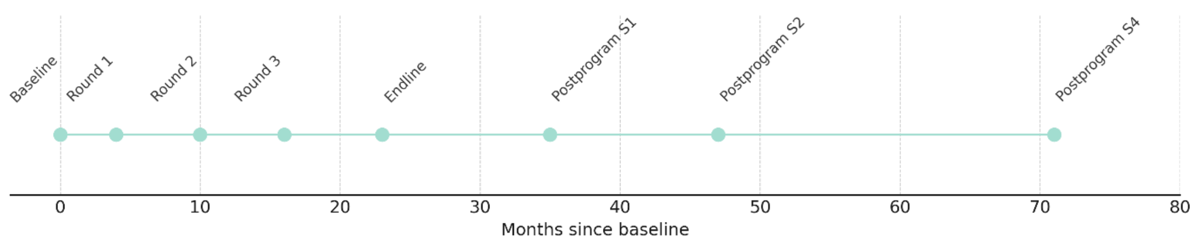
2.3 100Weeks' MEL system

This section assesses the quality of the data used by the internal monitoring of 100Weeks. We consider the following aspects:

- Alignment of indicators: Are the captured indicators aligned with the ToC?
- Data collection quality and process
- Data cleaning procedures
- Attrition and completeness

The centerpiece of the 100weeks MEL system is its monitoring surveys. These surveys extensively monitor program participants through frequent and broad-scoped assessments. Participants are surveyed at program start (hereafter referred to as the baseline), at three intermediary time points within the 100-week timeframe (round one, two, and three), and after 100 weeks have passed (hereafter referred to as endline). 100Weeks retraces program participants in three consecutive post-program surveys, extending up to four years after the program Figure 2 visualizes the timeline and frequency of monitoring surveys throughout the program duration.

Figure 2: 100Weeks data collection schedule over time (months)



The primary participant feedback mechanism for 100Weeks is its internal monitoring, and particularly the surveys conducted during the intermediary time-points. This robust MEL system underscores 100Weeks' identity as a learning organization, as demonstrated by key program adaptations such as the introduction of Village Savings and Loans Associations in 2020 after their data revealed the importance of group effects.

The surveys are designed to align with the Theory of Change of 100Weeks, ensuring that each specified outcome outlined in the ToC is measured by the questions. They cover a wide range of indicators related to different domains of wellbeing (e.g., mental wellbeing, physical health), standard of living (e.g., housing conditions, electricity access), social capital (e.g., participation in social life), competencies (e.g., entrepreneurial competencies), and income and assets (e.g., asset ownership). The Multidimensional Poverty Index (MPI) comprises several indicators from the domains of health, education, and standard of living.

2.3.1 Alignment of indicators with the ToC

The indicators currently utilized by 100Weeks provide a comprehensive overview of program outcomes. By employing the MPI, a broad poverty metric, 100Weeks effectively captures the lived experiences of poverty in a more holistic way than income-based measures alone would. Additional indicators, such as those related to mental health and intra-household power dynamics, further enrich the assessment of overall program impact. **100Weeks uses validated and cross-nationally comparable indicators to track multidimensional poverty and related outcomes.**

Laterite's review showed that many of the indicators used are aligned with internationally recognized instruments, including those developed by the World Bank and other global institutions. For example, multidimensional poverty is measured using the MPI framework, while food insecurity and diet diversity are assessed through well-established scales such as the Household Dietary Diversity Score (HDDS) and the Household Food Insecurity Access Scale (HFIAS). This approach ensures that findings are both robust and comparable across contexts. In addition, the risk of recall bias is mitigated by using internationally recognized and validated indicators. Also, the time horizons used in recall questions seem appropriate for the topics covered.

Some dimensions of poverty—such as income, financial inclusion, social deprivation, and environmental vulnerabilities—are not fully covered within the MPI. To address these gaps, 100Weeks collects additional data, including metrics on financial access and social capital, to complement the MPI and ensure a more comprehensive assessment. Increases in income are measured in the survey through yes/no questions, but do not measure it explicitly in monetary terms.

While the current multidimensional poverty approach is robust and should be maintained, adding monetary measures of poverty would strengthen it. At present, income increases are captured using simple yes/no questions. Because increasing incomes via cash-transfers and savings is a core pathway in the ToC, including monetary measures of poverty would strengthen the monitoring and evaluation of the program in two key ways.

First, monetary measures of poverty would allow more direct testing of key mechanisms in the theory of change. For instance, the effects of the cash-transfer program on investment decisions and diversification of incomes.

Second, monetary measures of income enable the calculation of return on investment, cost-effectiveness, and cost-to-benefit ratio, which are key metrics that can support scaling decisions and demonstrate value to funders.

We recommend including a short, rapid consumption module at baseline and endline, and possibly during follow-up surveys. These rapid consumption modules typically asks about the quantity and value of food and non-food items consumed over a recent reference period (e.g., 7 or 14 days for food and 30 days for non-food items). This addition would complement, not replace, the existing multidimensional indicators by capturing actual living standards alongside the broader multidimensional poverty and deprivation metric that 100Weeks already tracks effectively.

2.3.2 Data collection

The enumerators who collect these surveys are selected based on pre-defined criteria outlined in 100Weeks' internal documentation[1]. These criteria include prior surveying experience and demonstrated cultural sensitivity toward respondents. All enumerators are trained and paid for their work by 100Weeks. These trainings covers an interview guide with instructions for every question, practical preparation on how to use the SurveyCTO software, and how to conduct mock interviews. Mitigating the risk of social desirability bias is integrated in both the training and phrasing of survey questions. All participants of the 100Weeks program are surveyed longitudinally.

An example of a risk of bias was found by 100Weeks in Ivory Coast. A baseline-endline comparison of education levels raised concerns about under-reporting of education levels at baseline. Investigation revealed that some individuals, due to a reluctance to speak French, may have underreported their education to avoid the expectation of doing so. This issue received increased attention during enumerator training.

The progression of data collection is tracked using a monitoring dashboard, allowing visualization of completion statistics by individual participant, by group of participants, or by enumerator. Survey meta-data, including duration and GPS data, is available. In this initial review of their processes, we observe well-documented procedures that are supportive of quality data collection at scale.

100Weeks' data cleaning procedures include monthly 'dashboard meetings', in which the head of M&E meets with program managers and looks at incoming data in Tableau to check for issues such as outliers. When using the data, we came across a few minor issues related to cleaning. We had to remove 16 duplicate observations for Uganda and remove some households with empty 100Weeks identifiers. We recommend developing and implementing reproducible data cleaning scripts and standardized checklists to ensure data cleaning consistency across all datasets.

We find no evidence of selective (survey) attrition between baseline and endline, with overall attrition estimated at 7.6% for the 2020-2022 period.

Our investigation focused on selective attrition patterns in baseline-endline surveys spanning the start of the programme until 100 weeks later. We explored whether any attrition was correlated with the covariates relevant for the heterogeneity analysis. We concentrated on the 2020-2022 period following discussions with 100Weeks, who advised against using earlier

[1] 100Weeks has shared their Country Manual and an interview guide with Laterite. The Country Manual is a guiding document for all activities that a 100Weeks country team is involved in, aimed at ensuring consistency and running operations efficiently. The document covers responsibilities of roles, practical checklists, and operating procedures, including those of the M&E department

years due to the inconsistent application of the multidimensional poverty indicator—our key metric of interest—before 2020. Prior to this period, missing MPI values for participants and lower data quality made reliable analysis challenging.

Based on Laterite's analysis of the pooled sample data, attrition in endline data collection averaged approximately 7.6% between baseline and endline measurements. Note that this represents survey attrition rather than program attrition. 100Weeks maintains very low program dropout rates and keeps separate records of participants who remained in the program but were not surveyed for the endline datasets.

While we did not assess attrition in follow-up surveys beyond the endline period—since some participant groups had not yet been surveyed at 3 or 4 years post-program—it is likely that such follow-up data collection experiences higher attrition rates, which is typical for longitudinal studies of this nature.

Key insights so far:

Strong MEL foundation: 100Weeks has established a solid foundation for data-driven monitoring and evaluation that effectively supports programme improvement. While there are opportunities for enhancement—particularly in consumption measurement and strengthening data cleaning protocols—the current system provides a credible foundation for evidence-based program refinement and organizational learning.

Theory of Change alignment: The MEL framework demonstrates strong alignment between indicators and the program's Theory of Change, combining internationally validated measures like the MPI framework with supplementary metrics that capture broader poverty and well-being dimensions.

Quality data collection: Data collection quality is maintained through systematic enumerator MEL framework training, monitoring dashboards, and well-documented procedures, though some minor data cleaning issues suggest room for improvement through more standardized, reproducible processes. The low selective attrition rate of 7.6% in baseline-endline comparisons indicates reliable data completeness for core program evaluation.

Monitoring feeds into programme adaptations: The MEL system feeds into program adaptations as shown by modifications like introducing Village Savings and Loans Associations in 2020.

Recommendation: include consumption measures through a rapid consumption module: This would strengthen the system's ability to capture material living standards more accurately.

2.4 External research documents

We reviewed nine research documents (Table 1) provided by 100Weeks which they have commissioned to external researchers. Seven of those are master's theses (two qualitative studies, five quantitative), one publication of a randomized controlled trial (RCT) in a peer-reviewed journal and a research report.

Table 1: List of external research reports reviewed in the evidence review

Report title and author	Type of document	Country
1. Studying the effects of multifaceted cash transfers on the living-income gap. (2022). The Cash Lab.	Research report	Ivory Coast
2. Exploring the Impact of an Unconditional Cash Transfer Program on Cognitive and Relational Well-Being - A Study of Female Beneficiaries in Rwanda. (2020). B. Kell.	Qualitative thesis	Rwanda
3. Cash Transfers Improve Economic Conditions and Reduce Maternal Stress in Rural Ivory Coast. (2024). S. Wolf, S Kembou, A. Ogan, K. Jasinska. *Journal of Child and Family Studies.*	Article in peer-reviewed journal, midline results (RCT)	Ivory Coast
4. Empowering Change: Revealing the Impact of Cash Transfer Programs on Women's Empowerment. (2024). M. Saleh.	Quantitative thesis	Ivory Coast
5. A Pathway Out of Extreme Poverty - A Case Study of the Unconditional Cash Transfer Scheme to Extremely Poor Women in Musanze, Rwanda (n.d.). J. M. Anne.	Qualitative thesis	Rwanda
6. The Relationship between Income, Social Support, Mental Wellbeing and Empowerment in Developing Countries - Evidence from the 100Weeks program in Rwanda. (n.d.). A. Langener.	Quantitative thesis	Rwanda
7. The impact of financial shocks on mental health - An econometric panel data analysis. (2020). W.H. Hijmissen-Bonhof.	Quantitative thesis	Rwanda
8. The Economic Impact of Unconditional Cash Transfers in Developing Countries: Evidence from the 100Weeks program in Rwanda. (2020). A. Langener (2020, thesis)	Quantitative thesis	Rwanda
9. There is more to it than cash - The importance of a multi-faceted cash transfer program for women's empowerment, an empirical assessment. (n.d.). L. Klunder.	Quantitative thesis	Rwanda

More external studies are underway, including a large randomized controlled trial (RCT) conducted in collaboration with Wageningen University & Research (WUR) and partners in Rwanda.

The research conducted covers a broad range of outcomes and pathways in the posited Theory of Change, from intra-household decision-making to income changes and wellbeing. The takeaway message for each research document relevant to HereWeGrow is listed in Table 2. Most findings align with the intended effects of the program as outlined in the program's ToC. For example, the cash transfer approach aims to improve empowerment (e.g. having a say in joint decisions of the household) and mental wellbeing. The evaluations confirm that psychological wellbeing and empowerment improved in relation to the program (Kell, 2020, Saleh, 2024, Anne, n.d., Langener, n.d., Klunder, n.d.).

Economic wellbeing is seen by participants as a starting point for other dimensions of wellbeing (Kell, 2020). Interestingly, women's authority over their own earnings can decrease (Saleh, 2024). Regarding income improvements: a quasi-experimental study in Ivory Coast on cocoa farmers (The Cash Lab, 2022, document 1) estimated that **annual income from farming went up by 552** USD (treated 1,627 USD, control 1,075 USD, +51%) for program participants. Monthly income from alternative off-farm sources went up by 36 USD (control 17 USD, treated 53 USD, +211%). Care was taken that cash transfers were not reported as income. This is the best estimation in the research documents of the effect of the 100Weeks program on income for a population relevant to HereWeGrow. The MPI decreased by 0.04, which was not significant.

The RCT (Wolf et. al., 2024, midline results, document 3) shows that program participants are less likely to experience food insecurity and more likely to save more money. Program participants saved approximately 74 EUR more in the first program year compared to a group receiving another treatment (VSLA-only, no cash). MPI scores showed a small but significant decrease of 0.04.

Table 2: Key takeaway messages from external research documents

Document	Outcomes assessed	Key take-away message
1. The Cash Lab. (2022).	Annual income from cocoa farming (not including transfers), savings (can include transfers), MPI score	The 100Weeks program is increasing* annual income from cocoa farming (+$552; treated $1627, control $1075), increasing monthly income from alternative income streams (+$36; control $17, treated $53) and increasing total (current) savings (+$349; treated $461, control $111). MPI score decreases by 0.04 (meaning: poverty decreases) but was not significant.
2. B. Kell. (2020).	Well-being (e.g. quality of relationships, level of happiness), material well-being (e.g. assets, how many meals, weekly earnings)	The 100Weeks program improves* psychological wellbeing. Participants see material well-being as a departure point for other dimensions of well-being.
3. Wolf et. al. (2024).	Economic well-being (e.g. MPI, assets index) maternal stress (perceived stress scale), food insecurity (e.g. did any household member go to sleep hungry due to a lack of food), educational engagement, educational aspirations for a child	Midline results (1 year after start) indicate that the 100Weeks program improved mothers' economic well-being (e.g. saved ~€74 more over the last year compared to a control group with VSLA-groups but without cash-transfers) and reduced maternal stress. MPI scores decreased by 0.04 (0.27 standard deviation); a small but significant amount. Food insecurity decreased. No statistically detectable impact on parental educational engagement or educational aspirations of children was found.
4. M. Saleh. (2024).	Economic empowerment (e.g. started a new income generating activity), bargaining power (e.g. influence on joint decisions)	The 100Weeks program boosted women's economic empowerment and bargaining power regarding joint decisions, but reduced women's autonomy over their own earnings.
5. J. M. Anne. (n.d.).	Social capital (e.g. personal relationships, networks), physical capital (e.g. land ownership)	The 100Weeks program contributes to women's social, human, natural, physical, and financial capitals and their sense of agency.
6. A. Langener. (n.d.).	Income (weekly financial diaries, averaged monthly) well-being (self-reported, likert-scale, e.g. feeling unhappy/depressed)	The 100Weeks program is positively impacting* income and well-being, as suggested by SEM-modelling.

7. W.H. Hijmissen-Bonhof. (2020).	Financial shocks (e.g. weekly income compared to average weekly income, based on weekly financial diaries), mental health (self-reported, likert-scale, e.g. feeling reasonably happy)	Positive financial shocks do increase* mental health (negative shocks vice versa). Fluctuations in income negatively impact those with above-average mental health at the start.
8. A. Langener. (2020).	Basic need expenditure (excluding irregular expenses such as health care, based on weekly financial diaries), investment (e.g. farm inputs, inputs for business), net earnings	The 100Weeks program led to* more basic need expenditure, higher investment and net earnings. For example, weekly basic need expenditure went up by +1880 RWF (less than $2); pre: 4441, post: 6321).
9. L. Klunder (n.d.).	Personal empowerment (self reported, likert-scale, e.g. feeling reasonably happy), relational empowerment (e.g. influence on joint decisions)	The 100Weeks program improves* personal and relational empowerment, but it does not significantly correlate with indicators of general empowerment in the wider local society (e.g. how gender roles are believed to be perceived by the wider community).

*Please note that this is suggested by the results and not proven to be causally related

The contributions of the research documents to understanding causal pathways in more detail are depicted in Figure 24 in the Annex.

The nine research documents are systematically scored in five domains (A-E) using 11 guiding questions presented in the Assessment Framework (see Table 3). The answers to the guiding questions are then consolidated and scored per domain. These scores are color-coded (1 (1), 2 (2), 3 (3), 4 (4) – and up to 5 (5) for the Maryland Scale only), generating a heat map of strengths and weaknesses of the available evidence.

The consolidated substantiation of these scores is presented in the Annex (Table 19).

In summary, we assess the following areas:

- **Domain A —** Clarity and articulation of the ToC. Is the ToC well-articulated and relevant for the 100Weeks model?
- **Domain B —** Assumptions and causal pathways. Are hypothesized causal pathways outlined and tested; are underlying assumptions tested?
- **Domain C —** Adaptation and learning. Are insights and recommendations given to refine the ToC, based on empirical evidence?
- **Domain D —** Rigor, with emphasis on rigor for causal claims. Is an appropriate design employed; are analytical methods transparently applied?
- **Domain E —** Maryland Scale. The documents are scored on the Maryland Scale which goes from 1 to 5. Only research designs adhering to the highest standards for generating causal evidence score high on this stringent scale.

Table 3 Assessment framework for evidence assessment

Theme	Guiding questions	Scoring (Domain A-D: 1-4; Domain E: 1-5)
A. Clarity and articulation of ToC	- Does the research report describe a ToC relevant for the 100Weeks model?[1] - Are the core intervention components (cash transfers, training, savings groups) and the hypothesized causal pathways explicitly identified?	1 (Poor): The ToC is not described or is very unclear. 2 (Fair): The ToC is mentioned but lacks detail and clarity. 3 (Good): The ToC is clearly described with most core components identified. 4 (Excellent): The ToC is articulated in detail with explicit identification of all core components and causal pathways.
B. Assumptions and causal pathways	- Does it explicitly link its measured outcomes to the ToC, addressing the hypothesized causal pathways? - To what extent does the study examine intermediate outcomes or mechanisms predicted by the ToC? - Does the report test the underlying assumptions of the ToC? - Are contextual factors that might influence these pathways effectively integrated into the analysis?	1 (Poor): There is little or no link between the measured outcomes and the ToC; assumptions are not tested. 2 (Fair): Some links between outcomes and the ToC are present, but intermediate outcomes and contextual factors are only superficially addressed. 3 (Good): The report makes clear links, examines several intermediate outcomes, and discusses key assumptions and contextual factors. 4 (Excellent): There is a comprehensive and critical assessment of the ToC, with robust testing of assumptions, detailed analysis of intermediate outcomes, and full integration of relevant contextual factors.
C. Adaptation and learning	- Does the research report or evaluation provide insights or recommendations for refining the ToC based on empirical evidence?	1 (Poor): No recommendations or learning points are provided. 2 (Fair): Limited insights or recommendations, without clear linkage to the ToC. 3 (Good): The report offers practical recommendations and learning points that relate to aspects of the ToC. 4 (Excellent): The evaluation provides in-depth insights and robust recommendations that directly inform refinements to the ToC and program design.
D. Rigor (for causality)	- Does the evaluation employ an appropriate design (experimental, quasi-experimental, or robust observational) to test its causal claims? - Are the sampling methods and data collection techniques adequately described and justified? - Are the analytical methods clearly explained and transparently applied, with limitations or potential biases openly discussed?	1 (Poor): Design is inappropriate or inadequately explained, with major gaps in sampling and analysis. 2 (Fair): Design and methods are described but with significant limitations that reduce confidence in causal claims. 3 (Good): The design is generally appropriate, with clear and justified sampling and analytical methods, though some limitations remain. 4 (Excellent): The research report employs a rigorous design with well-documented, robust methods that effectively address potential biases and support strong causal inferences.
E. Maryland Scientific Methods Scale[2]	- In parallel, each study will be assigned a score on the Maryland Scientific Methods Scale (1–5).	1 (Poor): Weak design (e.g., no valid counterfactual). 2 (Fair): Somewhat weak design (e.g. no valid counterfactual, but adequate control variables or matching techniques are used to compensate for this) 3 (Good): Moderately robust quasi-experimental design (e.g. before-after comparison with a valid counterfactual. Baseline equivalence is shown, or differences are effectively accounted for). 4 (Very good): Robust research (e.g. a design with quasi-randomness in treatment). 5 (Excellent): Very strong design (e.g., randomized controlled trial with all relevant arguments made for attribution, such as little to no contamination between treatment and control).

1 This question is slightly rephrased, compared to the Inception Report to allow for a more meaningful review. This question now refers to 'a ToC relevant to the 100Weeks model', instead of 'the ToC', as the ToC of 100Weeks is not explicitly articulated to a large extent yet and those research documents effectively fill in some of the gaps.

Large Language Models (LLMs)—primarily Gemini 2.5 Pro—were used to aid the processing of all information presented in the nine research documents. A prompt was developed to enable the LLM to generate preliminary answers to the guiding questions, including references to specific sections of the documents, thereby expediting manual information retrieval. All summaries and conclusions drawn are ours. Table 4 shows the results of the scoring.

Table 4: Scoring of research documents from evidence assessment

	A. Clarity and articulation of ToC (1-4):	B. Assumptions and causal pathways (1-4):	C. Adaptation and learning (1-4):	D. Rigor (1-4):	E. Maryland scale (1-5):
1. The Cash Lab. (2022).	1	1	2	3	2
2. B. Kell. (2020).	4	3	4	2	1
3. Wolf et. al. (2024).	4	4	2	4	5
4. M. Saleh (2024).	3	2	1	3	4
5. J. M. Anne (n.d.).	3	3	3	2	1
6. A. Langener (n.d.).	3	2	2	2	1
7. W.H. Hijmissen-Bonhof. (2020).	2	2	2	3	1
8. A. Langener (2020).	2	2	2	2	2
9. L. Klunder (n.d.).	3	2	2	3	2

A key weakness of the body of evidence is that it seldom provides findings based on counterfactual comparisons, thereby hampering the research's ability to investigate changes that are attributable to the 100Weeks program.

Some studies elaborate the ToC of 100Weeks and develop causal pathways (column A scores reasonable). However, most quantitative theses primarily use data collected by 100Weeks that are not specifically tailored to the proposed ToC, limiting their ability to rigorously test causal pathways (column B). Two qualitative theses and the RCT are exceptions to this. The research is sometimes feeding back actionable insights to 100Weeks (column C). Some studies conduct their analysis with good rigor, although only one impact evaluation meets the highest standards for causal inference (column E). This limits the certainty of most conclusions about program effects.

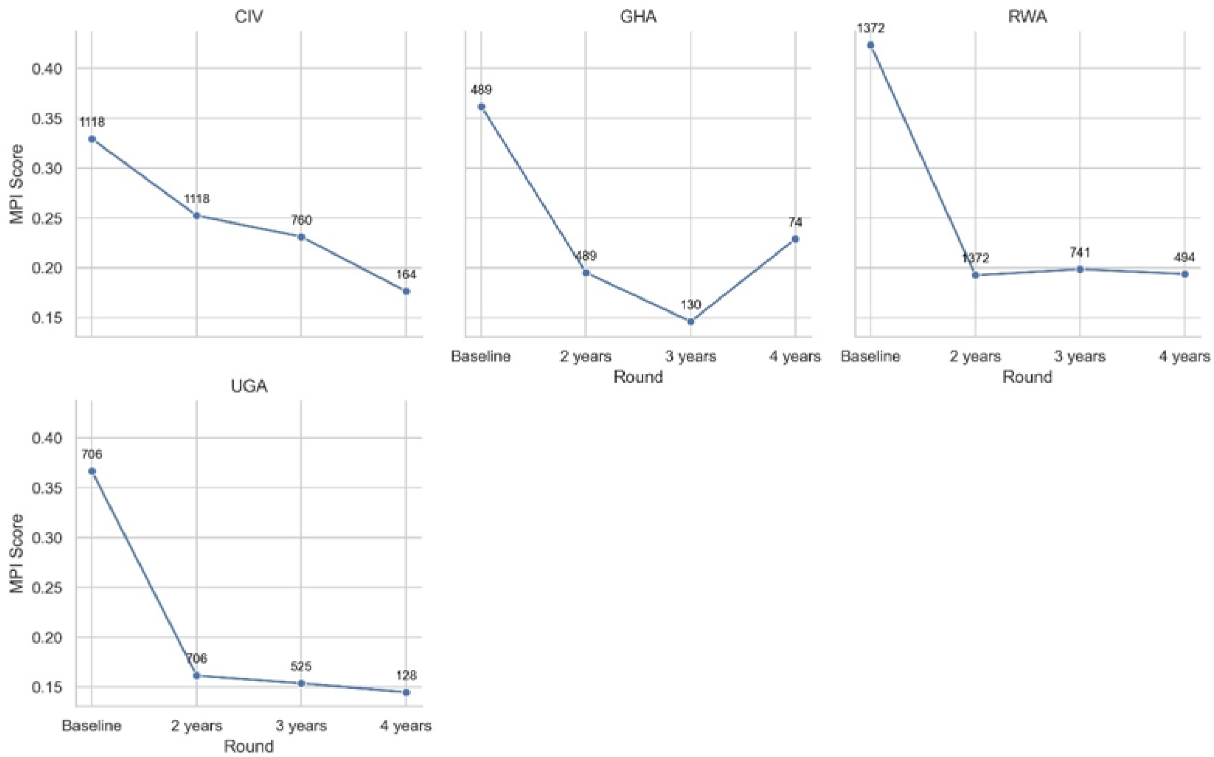
2.5 Observed program outcomes: Pre-post MPI changes

100Weeks shared all their monitoring data with Laterite. For this analysis, we included data of participants that had data both in baseline (pre) and endline (post) from Rwanda (N=1372), Ghana (N=490), Uganda (N=707), and the Ivory Coast (N=1118); giving a total of 3,687 program participants in this sample. Note that we distinguish between *MPI-scores* (a measure ranging from 0-1 where score 1 corresponds with high levels of multidimensional poverty) and *shares of participants classified as poor*. A participant with an MPI score above 0.33 can be classified as being poor (it is the equivalent of being fully deprived in one out of three poverty dimensions), which is the standard cut-off point

Using the current recommended MPI thresholds, **we find that 69% of participants are multidimensionally poor at baseline, which decreases to 17% of participants classified as multidimensionally poor after 100 weeks.** We observe an average decline in MPI scores from 0.38 at baseline to 0.21 at endline, a decrease of 0.17. **In Uganda the percentage of MPI-poor participants declines from 63% at baseline to 8% after 100weeks.** In Uganda the average decline is from 0.36 at baseline to 0.16 at endline (-0.20).[2]

The average decline in MPI scores is presented by country in Figure 3. The sample size is marked with every datapoint. Please note that the sample size is sufficient at baseline and endline, but drops significantly for 1 and 2 years after the program ends. Therefore, the primary interest is in the baseline-endline differences.

Figure 3 Average MPI score by round and country. Number of observations reported above.



For the subgroup that is MPI poor at baseline (69% of the total sample), poverty drops from 100% to 18% at endline. According to a claim presented on the website of 100Weeks[3], about 84% of the participants “stay out of poverty”. Our reanalysis is in line with 100Weeks’ findings.

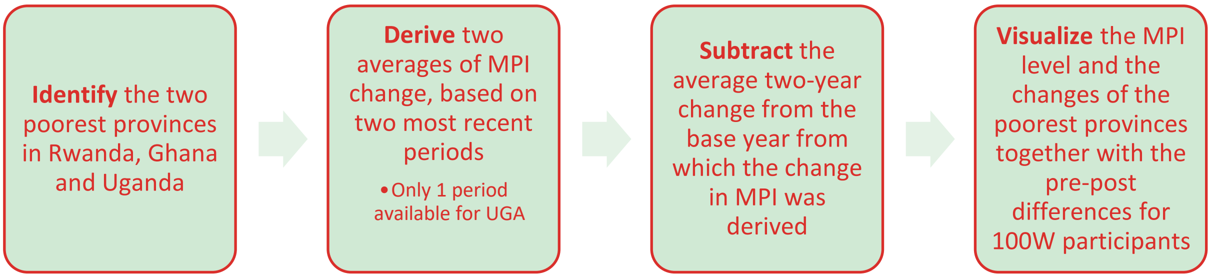
[2] More information about the MPI and its construction can be found here.
[3] https://100Weeks.org/impact

2.5.1 Contextualization of pre-post comparisons

To contextualize 100Weeks participants' poverty reductions, we compared changes in multidimensional poverty against national and regional trends in reference populations across Rwanda, Uganda and Ghana (data unavailable for Ivory Coast).

Using the Alkire-Foster method[4], we recalculated the 100Weeks MPI to match national and regional level statistics from OPHI's Global MPI database. Figure 4 shows the steps taken for this regional comparison.

Figure 4 Steps taken for regional pre-post comparisons



We focus on each country's two poorest regions to establish realistic comparison points for 100Weeks program participants (Figure 6). Ghana offers the strongest comparison with good temporal overlap. Rwanda provides acceptable overlap, while Uganda's outdated regional data limits comparison quality.

National trends in multidimensional poverty show consistent improvements across the targeted countries.

- Rwanda's MPI fell from 0.282 to 0.231 (2014-2020, -0.010 annually)
- Uganda's dropped from 0.349 to 0.281 (2011-2016, -0.014 annually)
- Ghana's MPI declined from 0.111 to 0.097 (2017-2022, -0.003 annually).
- Ivory Coast also shows declining poverty trends, though MPI data remains unavailable[5].

Participants in the 100Weeks program consistently started from higher poverty levels than regional comparison groups in Rwanda and Ghana (except for Ghana's 2021

[4] Following the Alkire-Foster methodology, the MPI is calculated by the formula H×A, in which H is the headcount ratio of poverty and A is the intensity of poverty among the poor, expressed as the average MPI for this subgroup. For example: if 40 percent of the population is poor – having an MPI of >0.33 – and the average MPI for this group is 0.5, MPI=H*A gives 0.4*0.5=0.2. For more information, see: https://www.unescwa.org/sites/default/files/event/materials/Alkire%20Foster%20Method%20and%20Global%20MPI.pdf

[5] https://data.worldbank.org/country/cote-divoire

baseline). This indicates that the program successfully targets comparatively poorer households.

Figure 6 Timeseries MPI (AF-method) 100W and regional comparison data

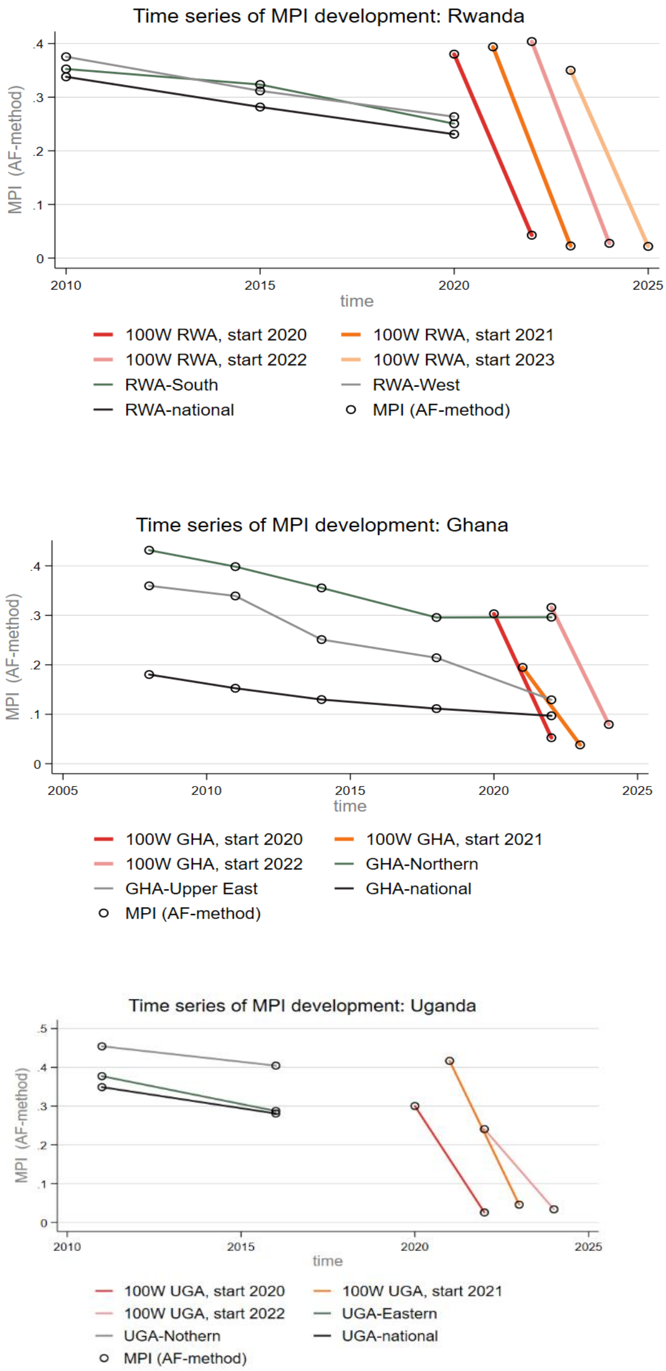
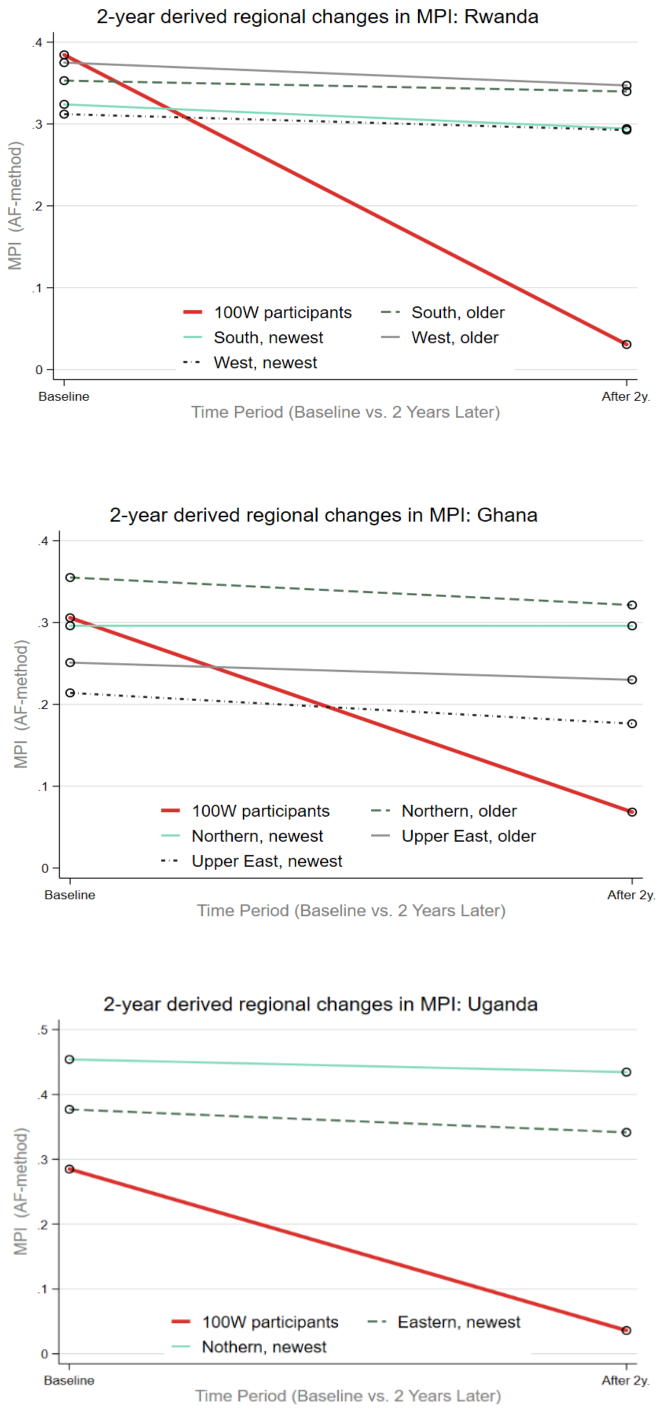


Figure 5 Derived 2-year change in MPI for program participants and regional reference populations



Next, we calculated typical two-year MPI changes for poor populations using regional data as comparison points. We derived annualized change scores from the available subnational data, which we then multiplied by two to match the relevant 100 weeks' (roughly 2 years) timeframe for the 100Weeks program. We then subtracted the MPI score that formed the base of the period[6].

We generated multiple comparison periods where data allowed. Rwanda and Ghana each provided two two-year periods per province, while Uganda's limited national and regional level data yielded only one period. Figure 5 visualizes these regional two-year changes alongside 100Weeks participants' improvements in MPI.

Program participants experienced dramatically faster poverty reduction than comparison populations. Participants' poverty declined 9 to 16 times faster than typical residents from the poorest provinces over the same timeframe:

- **Rwanda**: 100Weeks participants MPI dropped 0.3539 over two years, compared to 0.0226 in poor provinces (16x faster)
- **Ghana**: 100Weeks participants improved by 0.2374 versus 0.0231 regionally (10x faster)
- **Uganda**: 100Weeks participants declined by 0.2490 against 0.0276 regionally (9x faster)

We do note that data quality varies by country, affecting comparison strength. Rwanda (2010-2020 data) and Ghana (2008-2022 data) offer robust comparisons with recent regional MPI data and poverty levels similar to program participants. Uganda presents a weaker comparison due to outdated data (2011-2016 only) and less comparable initial poverty levels. Important timing limitations exist across all countries—for instance regional data in Uganda is from 2015 when poverty levels may have differed substantially from current conditions. Despite these temporal misalignments and data quality variations, all three countries consistently show the same pattern: program participants achieved substantially faster poverty reduction compared to regional comparisons.

2.6 Comparing research documents versus pre-post data

Interestingly, the pre-post measurements by 100Weeks indicate a substantially bigger drop in MPI scores (-0.17, from 0.38 to 0.21), compared to estimates presented in the external research documents (-0.04, estimated by both The Cash Lab (2022) and Wolf et. al., 2024). Several factors could account for this discrepancy.

Firstly, the RCT (Wolf et. al., 2024) is reporting midline results of a comparison of two treatments: a VSLA and Cash program, compared to a program with only VSLA. This is likely

[6] For example, in the Upper East province in Ghana, the MPI changed from 0.251 to 0.214 between 2014 and 2017/2018, and from 0.214 to 0.129 between 2017/2018 and 2022.The annual change is -0.0105 and -0.0189 for the two periods, respectively. To construct two benchmark lines for Upper East Province, 2*-0.0105 is subtracted from 0.251 and 2*-0.0189 is subtracted from 0.214, giving two lines that are presented in the figure.

an underestimation of the true program effects, which would be the effect of the 100Weeks program at endline, compared to receiving no program at all.

Secondly, the quasi-experimental research of The Cash Lab (2022) could be an underestimation of the typical program effect as well. This research was conducted on a group of cocoa farmers in the Ivory Coast, who are less deprived (wealthier) than the average program participant, leading to a smaller effect size as improvements in MPI are smaller for those who are richer. For the 100Weeks participants in the Ivory Coast, we observe a change in MPI of only -0.07 (from 0.32 to 0.25) in 100Weeks data, partly explaining the discrepancy between the estimations from the research documents and the pre-post changes in the data that we observe. Furthermore, the methodology of The Cash Lab (2022) limits the confidence in and precision of the effect size estimation. Their control group used for matching is formed at endline and no initial balance can be presented, limiting the confidence in attribution of observed income differences to the program. Their sample size is limited (N=375) and the estimated MPI change was not significant (no power calculation presented).

This means that the *smaller* estimations in the research documents can be reconciled with the *larger* pre-post differences observed. **Yet, since poverty shows a declining trend over the last decade in the program countries[7], pre-post comparisons by 100Weeks without a control group are likely an overestimation of the attributable program effect.** The national average MPI of Rwanda and Uganda is decreasing with about 0.01 per year.

2.7 Conclusion

The evidence review assessed 100 Weeks' program effectiveness through multiple lenses: it outlined the MEL procedures, the quality of the data generated, a re-analysis and contextualization of 100Weeks monitoring and evaluation data, and an assessment of externally commissioned research by 100Weeks.

100Weeks has a solid monitoring and evaluation system in place. The organization uses validated, internationally comparable indicators, including the MPI framework and established constructs like the Household Dietary Diversity Score. Importantly, 100Weeks employs multiple follow-up measurements beyond the standard baseline-endline comparison, tracking participants post-program completion. This longitudinal approach strengthens confidence in sustained impact, though sample sizes decline significantly in later periods.

Data collection procedures are well-documented with trained enumerators, systematic quality controls, and appropriate measures to mitigate social desirability and recall bias. **To ensure all relevant information is grasped, income or consumption could be included in the future**, as this would provide a quantifiable monetary measure of well-being, allowing for more precise comparisons across households, regions, or time periods. These metrics complement other indicators by capturing the material living standards in a way that is both interpretable and actionable. A short consumption module, often referred to as a rapid consumption module, typically asks about the quantity and value of food and non-food items consumed over a recent

[7] See 2.3.1, Contextualization of pre-post comparisons, for more details.

reference period (e.g., 7 or 14 days). This approach captures actual living standards more accurately than income alone, especially in informal or subsistence-based economies.

Our re-analysis of 100Weeks' monitoring data reveals substantial poverty reductions among program participants. Average MPI scores declined from 0.38 to 0.21 across all countries, representing a 0.17-point improvement. This translates to poverty rates dropping from 69% to 17% using the standard >0.33 poverty threshold. Uganda showed particularly strong results, with poverty declining from 63% to 8% of participants.

In contextualizing these findings and comparing multidimensional poverty against regional and national level reference populations we find strong indications of program effectiveness. Program participants achieved poverty reductions 9 to 16 times faster than typical poor populations in their regions. This finding holds consistently across Rwanda (16x faster), Ghana (10x faster), and Uganda (9x faster), despite data quality limitations in the reference data in some countries. Participants also started from higher poverty levels than regional comparison groups, indicating effective targeting of more vulnerable households.

Nine external research documents provide additional perspectives on program effectiveness. **The research confirms key program pathways: improvements in empowerment, mental well-being, savings behavior, and income diversification**. Notably, one quasi-experimental study found annual farming income increased by 552 USD (+51%) and alternative income sources by 36 USD monthly (+211%). Of special interest to HereWeGrow are the observed increases in savings, investments in business and farm inputs, and discovering new income streams. The research evidence shows mixed methodological rigor.

Only one study (Wolf et al., 2024) meets the highest standards for causal inference (Maryland Scale 5). Most quantitative studies rely on 100Weeks' data without tailored research designs, limiting their ability to rigorously test causal pathways. The findings in most studies likely overestimate effects as, without solid counterfactual designs, they are likely to combine both secular trends in, for instance, multidimensional poverty declines, and the effect of the program itself. Two qualitative studies and the RCT provide the strongest evidence for understanding program mechanisms and impacts. **A key area for improvement in future commissioned research is to focus on generating counterfactual-based evidence so that changes observed can be attributed to the program.**

The convergence of evidence from multiple sources strongly suggests meaningful program impact. For example, participants in the 100Weeks program show a substantial reduction in multidimensional poverty that markedly exceeds trends in regional reference populations. It remains uncertain what the precise magnitude of program-attributable effects are as most studies do not rely on counterfactual designs.

All in all, the available evidence indicates that the 100Weeks model is an effective poverty reduction intervention. However, expanding the evidence base with more rigorous impact evaluations that incorporate counterfactual designs is essential to reliably quantify the program's causal net effects.

3 Heterogeneity analysis of MPI score changes between baseline and endline

3.1 Objective

The heterogeneity analysis aims to investigate which recipients experience the biggest benefits from the program. By examining how outcomes vary across different subgroups (such as type of participant, age, number of household members, partner status, MPI poverty at baseline, literacy of household head, debt outstanding at baseline, land ownership at baseline, start year of the program), we can identify heterogeneity in change of MPI scores for these groups. We will focus on treatment group data (baseline and endline) to assess heterogeneous impacts on the key outcome, the Multidimensional Poverty Index (MPI). We examine the contribution of each component to poverty reductions.

3.2 Methodology and data

3.2.1 Methodology

For the main outcome of interest (MPI) we estimate that the change of MPI scores between baseline (BL) and endline (EL) is 0.17 points, which amounts to a 52% decrease in the proportion classified as poor. It is important to highlight that this is a solely pre-post effect, not an attributable impact, since there is no control group in the study. This is rather the observed change in the outcome for program participants. This average change estimate conceals variation across different subgroups. We apply several methods to explore the heterogeneity in the MPI score decrease between baseline and endline.

We first use **Random Forests**, a standard machine learning algorithm, to explore which variables are most important in explaining variation in changes to MPI scores between baseline and endline. Random Forests works by building an ensemble of decision trees, each trained on a random sample of the data (with replacement). At each split in a tree, the algorithm considers a random subset of variables to reduce overfitting and increase model diversity. The final prediction is an average across all trees. Although Random Forests are typically used for prediction, they can also be used for variable importance analysis. In our case, we use them to identify which features are most strongly associated with changes in MPI scores**. These variables are then used to guide subsequent regression analyses, where we examine how MPI change varies across subgroups in terms of direction and magnitude.** We then construct coefficient plots that illustrate the regression results. As explanatory variables, we avoid including any of the individual level items used in the construction of the MPI score; including housing characteristics, education of household head, and assets. We used covariates measured at baseline for the heterogeneity analysis. All results are presented from regressions that control for all covariates listed.

We include the characteristics listed in Table 5. We determine whether there are notable differences in the change in MPI between BL and EL based on the set of variables listed.

Table 5: Covariates (measured at baseline) included in the heterogeneity analysis

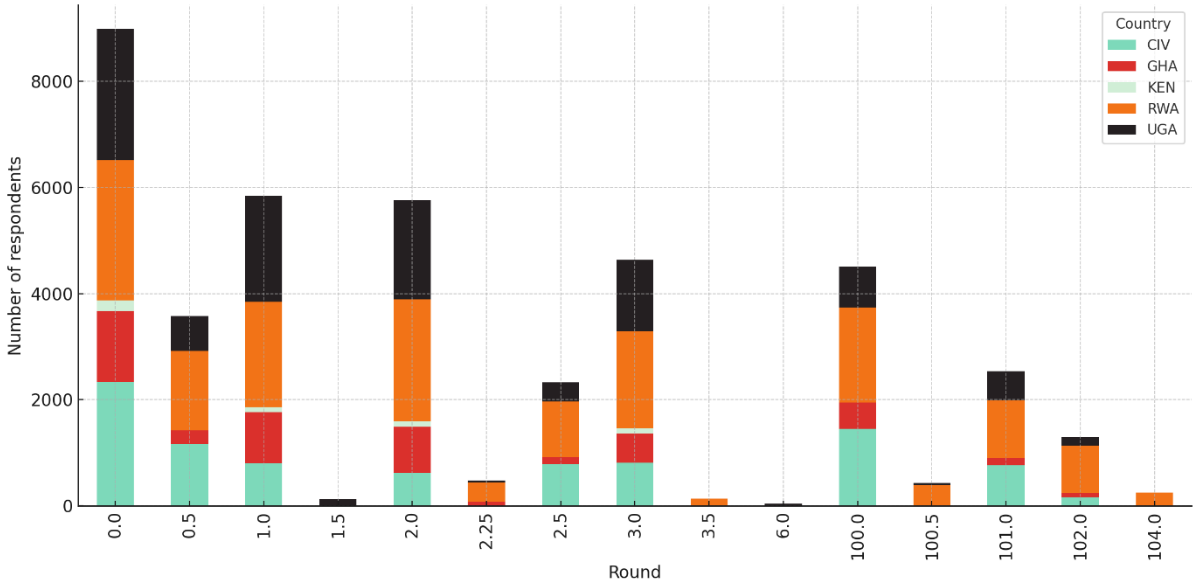
Variable	Description
Age of household head	Continuous
Country	Uganda (reference), Rwanda, Ivory Coast, Ghana
Multidimensional Poverty Index (MPI) poor in 4 categories	Not poor, vulnerable to poverty, poor, severely poor
Household size	Continuous
Owns land	Yes, owns 1 piece of land (reference) No, does not own any land Yes, owns >1 piece of land
Debt	Yes, No
Partner	Spouse in household (reference), partner in HH, spouse not in HH, partner not in HH, partner died, no partner
Start year	Year of start of 100Weeks program 2020 (reference) to 2022
Home ownership	Owns (reference), rents, caretaker
Type of farmer	Subsistence, conservation, coffee-farmer, cocoa-farmer
Dummies for group	100 Weeks group specific dummies (hundreds of groups of usually 20 participants)

3.2.2 Dataset description

For the assignment, 100Weeks provided Laterite with a complete dataset covering five countries: Ivory Coast (CIV), Ghana (GHA), Kenya (KEN), Rwanda (RWA), and Uganda (UGA). Figure 7 shows the distribution of the data by country and survey round. The survey follows participants over time, starting with a **baseline** (Round 0), conducted at the very beginning of the program. Subsequent rounds are conducted monthly or at key intervals throughout the 100-week program and beyond. Each round is indexed by the number of months since the baseline.

The dataset contains information on other points in time that were reported for specific countries. For the purpose of this review, in Figure 7 below, 0 corresponds to baseline, 100 to endline and the integers correspond to the schedule presented in Figure 2. Non-integer values correspond to surveys carried out for some countries and in varying magnitudes of number of participants.

Figure 7 Number of observations per country and round in 100Weeks raw dataset



For the analysis, we drop Kenya from the dataset, as well as focus only on groups that started after 2020. This is because the implementation of the full MPI score was only applied fully and consistently for groups after this start year. **We only focus on BL (round 0) and EL (round 100) comparisons, which corresponds to close to two years after the start of the program.**

To prepare the dataset for the heterogeneity analysis, some additional steps were taken. This included a) removing 11 duplicate observations from Uganda and observations with empty unique household identifiers as previously mentioned, b) creating a time variable (start year) based on the time stamp of the survey to be used to study period effects given the presence of exogenous events within this time period (such as the COVID-19 pandemic in 2021).

3.2.3 Summary statistics of variables of interest

To further explore the data, we constructed summary statistics for the MPI score, as well as the explanatory variables that are used in the heterogeneity analysis. Table 6 shows the mean and standard deviation, as well as the proportions for relevant variables for participants that will be included in the analysis. This excludes missing in either MPI score or explanatory variables.

Table 6: Summary statistics by round

	Round 0 (BL)	Round 100 (EL)	p-value
	N=3685	N=3685	
Country			1.00
CIV	1118 (30.3%)	1118 (30.3%)	
GHA	490 (13.3%)	490 (13.3%)	
RWA	1372 (37.2%)	1372 (37.2%)	
UGA	707 (19.2%)	707 (19.2%)	
Partner of household head			<0.01

	Round 0 (BL)	Round 100 (EL)	p-value
Spouse in HH	2309 (63.0%)	2466 (66.9%)	
Partner in HH	494 (13.5%)	436 (11.8%)	
Spouse not in HH	90 (2.5%)	66 (1.8%)	
Partner not in HH	114 (3.1%)	76 (2.1%)	
Partner died	179 (4.9%)	212 (5.8%)	
No partner	481 (13.1%)	430 (11.7%)	
Gender of household head			0.72
Man	98 (2.7%)	93 (2.6%)	
Woman	3479 (97.3%)	3477 (97.4%)	
Number of household members	4 (2)	4 (2)	0.24
Age categories			<0.01
<20	35 (0.9%)	16 (0.4%)	
20-29	1066 (28.9%)	807 (21.9%)	
30-39	1649 (44.7%)	1615 (43.8%)	
40-49	611 (16.6%)	863 (23.4%)	
50-59	235 (6.4%)	266 (7.2%)	
60+	91 (2.5%)	120 (3.3%)	
Home ownership			<0.01
House - rented	327 (8.9%)	352 (9.5%)	
House - caretaker	425 (11.5%)	272 (7.4%)	
House - owned	2935 (79.6%)	3063 (83.1%)	
Literacy of household head			0.02
Literacy - yes	2259 (61.3%)	2353 (63.9%)	
Literacy - no	1428 (38.7%)	1329 (36.1%)	
Land ownership status			<0.01
Yes, owns one piece of land	1773 (48.9%)	221 (31.8%)	
No, does not own land	898 (24.8%)	106 (15.2%)	
Yes, owns >1 piece of land	956 (26.4%)	369 (53.0%)	
Has debt outstanding			<0.01
Debt - yes	2022 (56.0%)	1754 (47.6%)	
Debt - no	1586 (44.0%)	1933 (52.4%)	
MPI score (lower is less poor)	0.38 (0.15)	0.21 (0.12)	<0.01
MPI poverty (4 cat)			<0.01
MPI non-poor	480 (13.0%)	2,085 (56.6%)	
MPI non-poor but vulnerable	657 (17.8%)	986 (26.8%)	
MPI poor	1682 (45.6%)	477 (12.9%)	
MPI severely poor	866 (23.5%)	137 (3.7%)	
MPI poverty (binary)	2548 (69.1%)	614 (16.7%)	<0.01
Target group			
Cocoa farmers	1434 (38.9%)	1434 (38.9%)	1.00
Coffee farmers* (4 EUR per week for 50 weeks)	86 (2.3%)	86 (2.3%)	
Conservation	191 (5.2%)	191 (5.2%)	
Subsistence farmers	1974 (53.6%)	1974 (53.6%)	

3.2.4 MPI score change between endline and baseline

We begin by exploring the distribution of the outcome of interest: the change of MPI between baseline and endline (100 weeks after BL). The figure below shows the distribution of MPI score difference between BL and EL, with the red line at 0.

Figure 8 Distribution of MPI score difference between BL and EL. Red line at 0

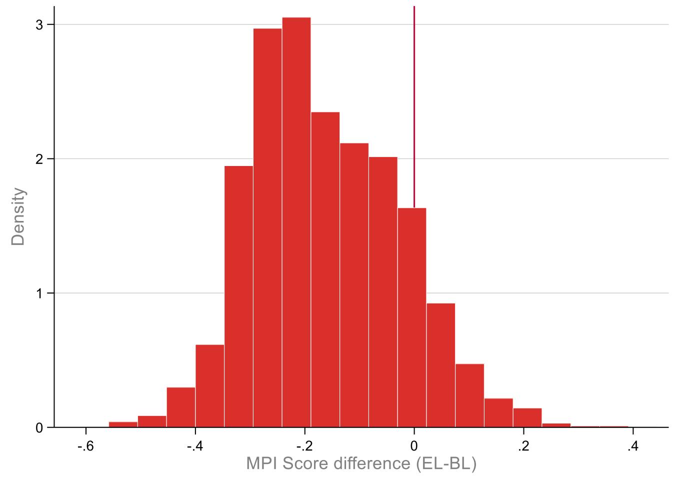


Figure 8 shows the distribution of the difference between MPI in the two time periods. There is a group of participants that have increasing MPI scores (have become poorer), albeit a small proportion. The number of respondents that experienced no change in poverty status (using the 0.33 threshold) is 70%, followed by those who graduated from poverty (28%) and those who fell deeper into poverty (2%) between BL and EL. The heterogeneity of change in the outcome of interest can be also visualized if we divide the outcome in quintiles by magnitude. Quintile 1 is the one with the most decrease in poverty (e.g negative change in MPI score), and quintile 5 is the one with the lowest decrease (or even increase).

Figure 9: MPI score change by quintile

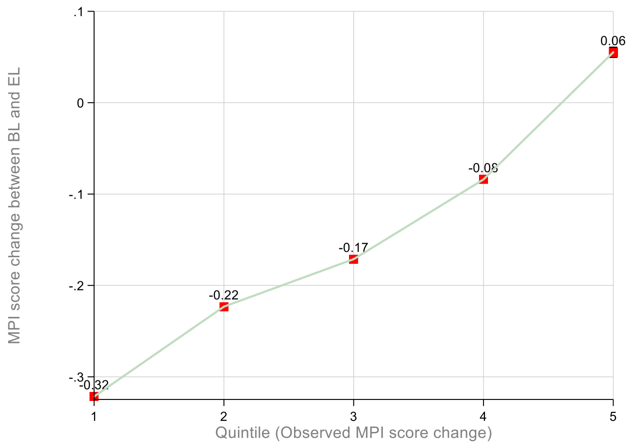


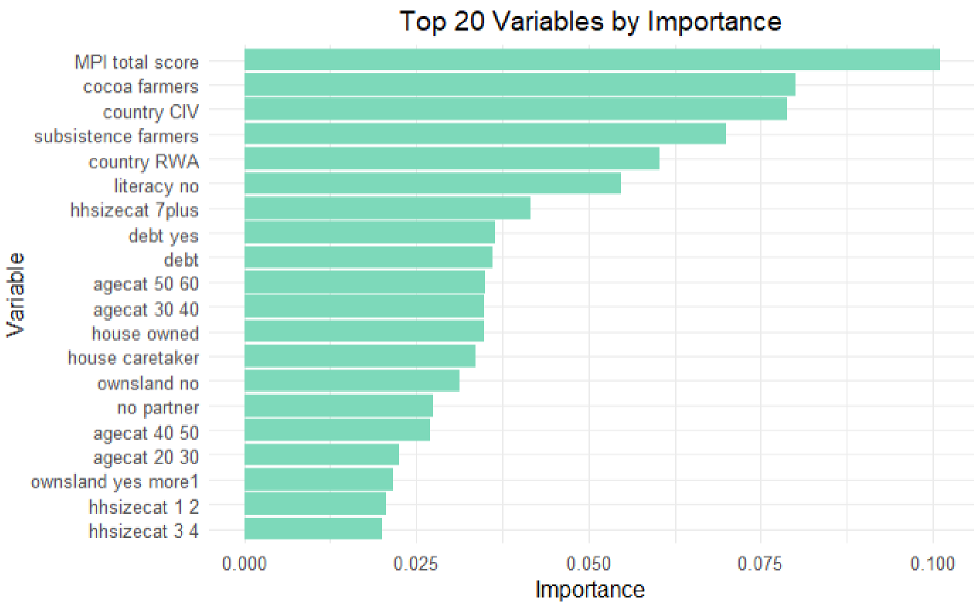
Figure 9 shows the average observed change in MPI score between baseline and endline across quintiles, ranked from the lowest to the highest observed impact. Each dot represents the mean change in MPI within each quintile. We observe a clear monotonic gradient: individuals in the lowest quintile (Q1) experienced the largest reduction in MPI scores (–0.35), while those in the highest quintile (Q5) saw a small increase in MPI score (+0.04). This suggests heterogeneity in outcomes across subgroups. We further explore which household characteristics are associated with larger versus smaller improvements in MPI.

3.3 Results

3.3.1 Overall heterogeneity

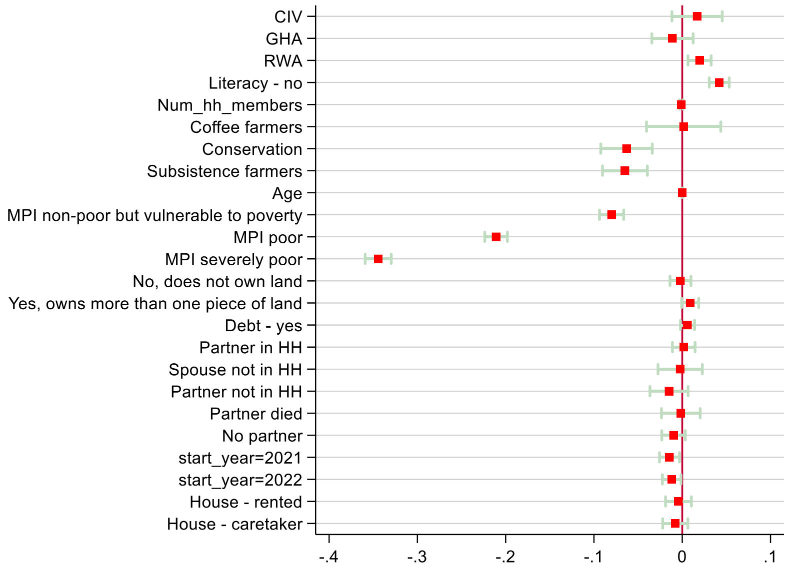
Figure 10 shows the variable importance derived from the Random Forests modelling. It shows that the MPI score at baseline, country, literacy of household head, household size, target group (cocoa, coffee or subsistence farmer), age of household head, home ownership, land ownership, start year, and debt outstanding; are the most important predictors included in the RF modelling using MPI difference between baseline and endline as the outcome. Cohort-level effects (group number in the context of 100 Weeks) appear not to be important as per the model.

Figure 10: Variable importance derived from Random Forest model for 20 variables



We then use regression analysis to explore how the covariates of highest importance are correlated to variation in changes in MPI between rounds. We used OLS models with robust standard errors. We exclude fixed or random effects for group, as they do not appear as important predictors in the Random Forests modelling. Figure 11 presents the coefficient plot for the regression analysis.

Figure 11: Coefficient plot from regression using MPI change between EL and BL as dependent variable.



Graph omits reference categories for categorical variables. All independent variables are measured at BL. Horizontal axis is change in MPI score between BL and EL.

Figure 11 shows the coefficient output from the regression described above, the horizontal axis is the change in MPI score between baseline and endline. **The graph shows that by far, the most important decreases in MPI score between BL and EL are associated with initial levels of poverty at baseline.** There is a clear gradient, with the severely poor at BL gaining the most in MPI score at baseline, followed by the poor and to a lesser extent the vulnerable to poverty relative to those that are not poor at BL.

The graph indicates that relative to cocoa farmers, conservation and subsistence farmers had a greater reduction in MPI score. Coffee farmers are not significantly different than cocoa farmers. The graph shows that Rwandan participants were associated with a slight increase in MPI score relative to Ugandan participants. Households where the head is not literate have a larger, statistically significant increase in MPI score relative to those that are literate. Households that own more than one piece of land at baseline is associated with an increase in MPI score at endline relative to those who own only one piece of land. Similarly, households that have no partner at baseline are associated with a greater decrease in MPI score relative to households that have the spouse in the household. As cocoa farmers are only present in Ivory Coast this means that we cannot distinguish the effect of targeting cocoa farmers versus targeting participants in Ivory Coast. However, the key results that is: large MPI reductions in Uganda and Rwanda and smaller reductions in Ghana and Ivory coast are robust to removing the cocoa farmer categories.

These results suggest that:

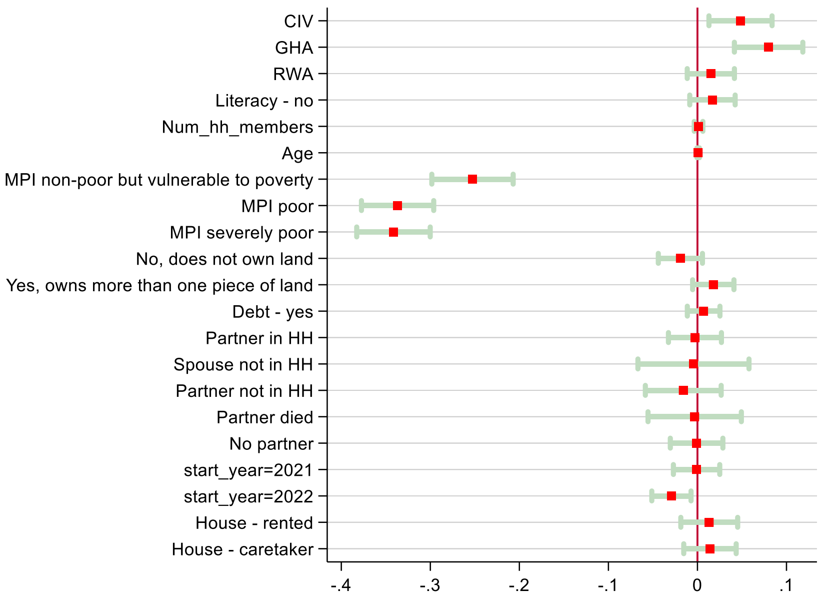
1. Households that were poorer at baseline saw much bigger improvements in their MPI scores by endline compared to those who were not poor at the start.

2. There are country differences in the change in MPI score between baseline and endline. Rwanda performs slightly worse than Uganda in terms of MPI decrease between the two periods (larger score means poorer households).
3. Subsistence agriculture and conservation farmers had a greater decrease in MPI score relative to cocoa farmers and coffee farmers. A direct comparison is for coffee farmers is not possible with other groups since coffee farmers are also receiving a different intensity of the cash transfer. This means for a shorter time – 50 weeks, and a reduced amount – 4 euros per week, translating to roughly 25% of the total amount transferred to the other groups.
4. Household heads who are not literate are associated with an increase in MPI score (higher score is poorer household). This suggests illiterate households might require additional support to benefit fully from the program. Starting the program in 2022 is associated with a decrease in MPI score.

3.3.2 Increasing MPI scores

Next, it is interesting to explore the group that experiences an increase in MPI score (i.e. the group of people that experiences an effect in the opposite direction i.e. multidimensional poverty increases). We dichotomize this to explore only those who have an increase in MPI score between BL and EL (439 participants or 12% of the total). Figure 12 below shows the results of a linear probability model that uses as dependent variable whether a participant has a positive MPI score change between baseline and endline (i.e. has become poorer). The horizontal axis can be interpreted as a change in probability.

Figure 12 Coefficient plot for linear probability model using as dependent variable whether a participant has a positive MPI score change (i.e. becomes poorer) between EL and BL.



All covariates Graph omits reference categories for categorical variables. All independent variables are measured at BL. Horizontal axis is a change in probability.

Figure 12 shows that households in the poorest baseline categories of the Multidimensional Poverty Index (MPI) are 35% less likely to become poorer between the start and end of the program. The data suggests that participants in Ghana and being in the richest group at baseline, are more likely to fall into the subgroup with rising MPI scores, an outcome that differs from the program's overall average effect. This subgroup corresponds to the top quintile in terms of MPI score change across rounds, with an average increase of 0.04 in MPI. This is only a modest increase. Only a small number of participants experienced substantial rises in MPI scores, indicating a shift toward greater poverty.

3.3.3 MPI disaggregated

Finally, it is of interest to explore what specific components of the MPI are changing over the two periods. Figure 13 and Table 7 show whether a participant experienced an improvement remained the same as in BL, or worsened in the outcome (MPI for the domain was 0 and became 1 – e.g. became at risk for poverty for that domain).

Figure 13 MPI changes (improved, no change, worsened by country and domain)

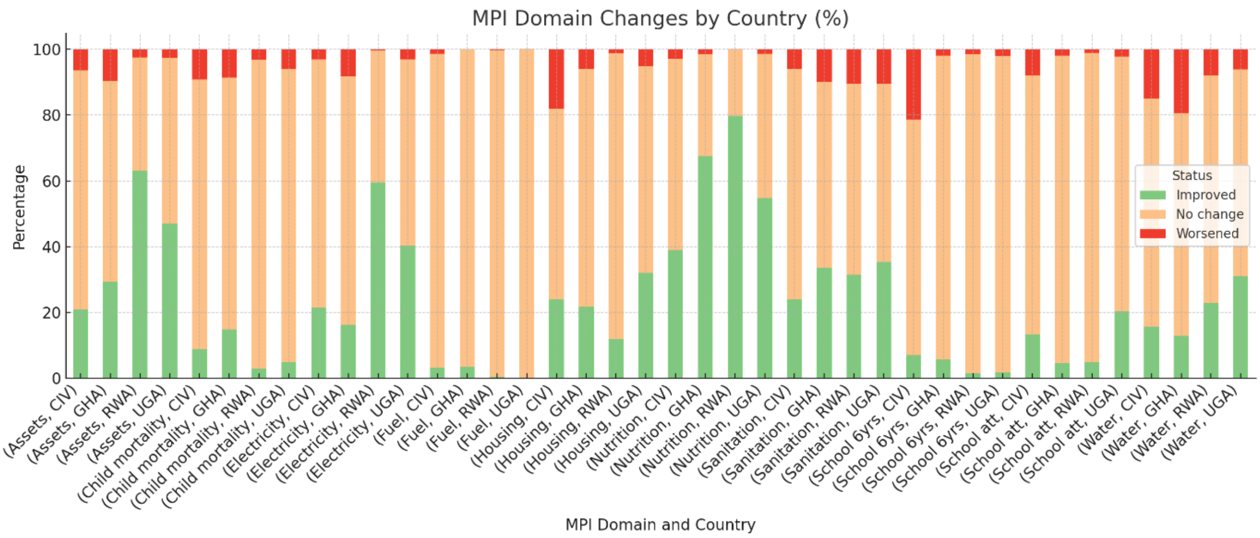


Table 7 Change in MPI by component and country between baseline and endline

	CIV	GHA	RWA	UGA	p-value
	N=1,118	N=489	N=1,372	N=706	
Child mortality					<0.001
Improved	98 (8.8%)	72 (14.7%)	40 (2.9%)	34 (4.8%)	
No change	917 (82.0%)	375 (76.7%)	1,289 (94.0%)	630 (89.2%)	
Worsened	103 (9.2%)	42 (8.6%)	43 (3.1%)	42 (5.9%)	
Nutrition					<0.001
Improved	436 (39.0%)	330 (67.5%)	1,094 (79.7%)	387 (54.8%)	
No change	650 (58.1%)	152 (31.1%)	276 (20.1%)	310 (43.9%)	
Worsened	32 (2.9%)	7 (1.4%)	2 (0.1%)	9 (1.3%)	

	CIV	GHA	RWA	UGA	p-value
Fuel					<0.001
Improved	36 (3.2%)	16 (3.3%)	6 (0.4%)	0 (0.0%)	
No change	1,067 (95.4%)	473 (96.7%)	1,363 (99.3%)	705 (99.9%)	
Worsened	15 (1.3%)	0 (0.0%)	3 (0.2%)	1 (0.1%)	
Electricity					<0.001
Improved	238 (21.3%)	79 (16.2%)	817 (59.5%)	284 (40.2%)	
No change	846 (75.7%)	370 (75.7%)	551 (40.2%)	401 (56.8%)	
Worsened	34 (3.0%)	40 (8.2%)	4 (0.3%)	21 (3.0%)	
Water					<0.001
Improved	174 (15.6%)	63 (12.9%)	314 (22.9%)	219 (31.0%)	
No change	777 (69.5%)	331 (67.7%)	949 (69.2%)	444 (62.9%)	
Worsened	167 (14.9%)	95 (19.4%)	109 (7.9%)	43 (6.1%)	
Sanitation					<0.001
Improved	267 (23.9%)	164 (33.5%)	431 (31.4%)	249 (35.3%)	
No change	784 (70.1%)	277 (56.6%)	797 (58.1%)	383 (54.2%)	
Worsened	67 (6.0%)	48 (9.8%)	144 (10.5%)	74 (10.5%)	
Housing					<0.001
Improved	266 (23.8%)	106 (21.7%)	162 (11.8%)	226 (32.0%)	
No change	649 (58.1%)	354 (72.4%)	1,193 (87.0%)	444 (62.9%)	
Worsened	203 (18.2%)	29 (5.9%)	17 (1.2%)	36 (5.1%)	
Assets					<0.001
Improved	232 (20.8%)	143 (29.2%)	864 (63.0%)	331 (46.9%)	
No change	815 (72.9%)	299 (61.1%)	475 (34.6%)	357 (50.6%)	
Worsened	71 (6.4%)	47 (9.6%)	33 (2.4%)	18 (2.5%)	
School attendance (children)					<0.001
Improved	148 (13.2%)	22 (4.5%)	67 (4.9%)	143 (20.3%)	
No change	882 (78.9%)	458 (93.7%)	1,290 (94.0%)	548 (77.6%)	
Worsened	88 (7.9%)	9 (1.8%)	15 (1.1%)	15 (2.1%)	
School 6 years (all HH members)					<0.001
Improved	77 (6.9%)	28 (5.7%)	18 (1.3%)	12 (1.7%)	
No change	802 (71.7%)	452 (92.4%)	1,334 (97.2%)	680 (96.3%)	
Worsened	239 (21.4%)	9 (1.8%)	20 (1.5%)	14 (2.0%)	

The results of the analysis reveal significant variation across the four countries (Ivory Coast, Ghana, Rwanda, and Uganda) with respect to the changes in various MPI domains.

Child Mortality: A large proportion of respondents in all countries experienced "no change" (82%-94%), with the highest improvement in Ghana (14.7%) and the lowest in Uganda (4.8%). However, the proportion of worsened cases remained relatively consistent across countries, ranging from 2.9% in Rwanda to 9.2% in Ivory Coast. These differences are statistically significant ($p < 0.001$).

Nutrition (HFIAS): Significant improvement was observed in Rwanda (79.7%) and Ghana (67.6%), while the lowest improvement was in Uganda (54.9%). The proportion of "no change" was highest in Ghana (31.0%) and lowest in Rwanda (20.1%), with significant country-based differences ($p < 0.001$).

Fuel: Nearly all respondents reported "no change" in fuel access across the countries, with Rwanda and Uganda having nearly universal reports of no change (99.3% and 99.9%, respectively). A small fraction reported improvement or worsening, with the lowest rates of improvement found in Uganda (0%).

Electricity: Rwanda exhibited the highest improvement rate (59.5%), while Uganda showed a slightly lower rate (40.3%). The "no change" category was the highest in Ghana (75.7%), while worsening electricity access was most notable in Ghana and Uganda.

Water: There was a substantial proportion of respondents in Uganda (31.1%) who reported improvement in water access, while Rwanda (22.9%) and Ghana (12.9%) had lower improvement rates. The "worsened" category was most significant in Ivory Coast (14.9%) and Ghana (19.4%).

Sanitation: The proportion of improvement was highest in Uganda (35.4%), with Ghana showing the lowest (33.5%). A higher percentage of "worsened" cases was seen in Rwanda (10.5%), whereas Ivory Coast had a relatively lower percentage (6.0%).

Housing: Housing access showed the highest improvement in Uganda (32.1%), while the lowest was in Rwanda (11.8%). The "no change" category was particularly high in Rwanda (87.0%), indicating little progress in this area.

Assets: Rwanda exhibited the highest rate of improvement in asset ownership (63.0%), while Uganda had the lowest (46.8%). The proportion of worsening asset access was higher in Ghana (9.6%) and Ivory Coast (6.4%).

Across all domains, the differences between countries are statistically significant ($p < 0.001$), indicating that improvements and regressions in the MPI categories vary widely across countries.

A different way of exploring the relative importance of the different components of MPI is to calculate the contribution of each of the components to the observed change in MPI score for those who exited poverty between BL and EL. Figure 14 show this calculation by country.

Figure 14 Attribution (%) of each MPI score component to the decreases in MPI score between EL and BL by country. Only for those who exited poverty between BL and EL.

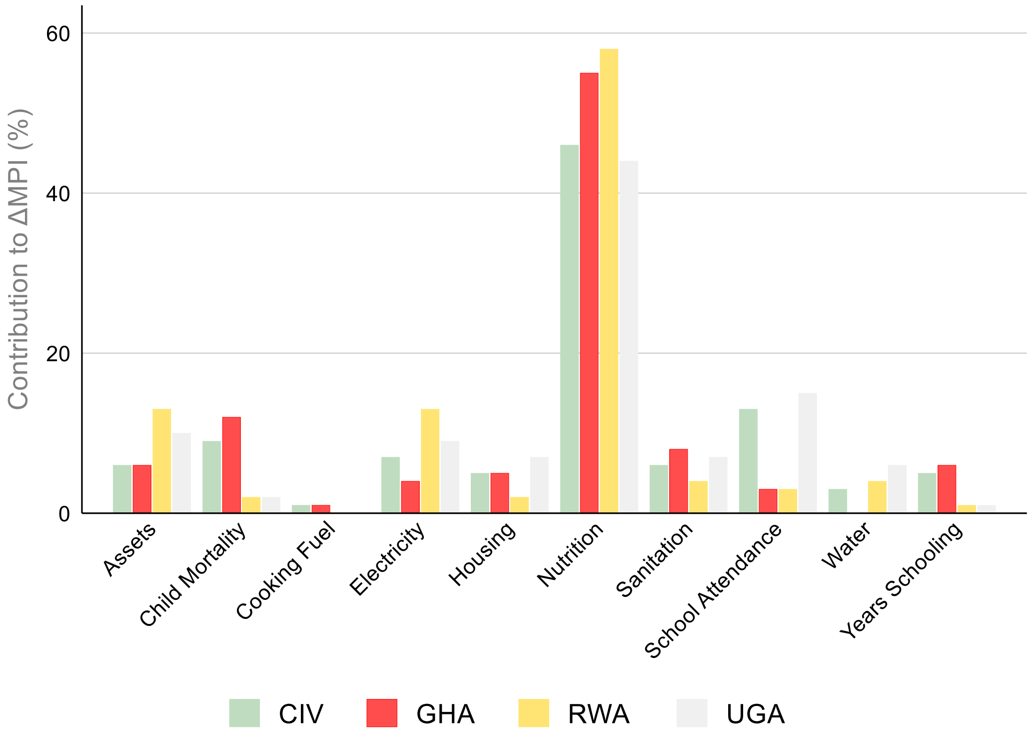


Figure 14 shows that by far, the biggest contributor to MPI score change for those who exited poverty among all participants is nutrition. Among these, **Nutrition** stands out as the most significant driver, accounting for **over half (52.5%)** of the total reduction in poverty. This is followed by **Assets (10.4%)**, **Electricity (9.9%)**, and **School Attendance (7.2%)**, which also play substantial roles. Mid-level contributors include **Sanitation (5.6%)**, **Child Mortality (4.3%)**, **Housing (4.0%)**, and **Water (3.8%)**. Meanwhile, **Years of Schooling (2.1%)** and **Cooking Fuel (0.3%)** contribute minimally.

For all countries, nutrition remains the most important contributor. There is some heterogeneity in the rest of the factors, with variation within assets, child mortality, electricity and school attendance between countries. Cooking fuel contributes minimally as most participants are classified as poor in both BL and EL in this component.

Key insights in chapter 3:

Poverty reduction is strongest for those who are multidimensionally poor at baseline
There is a clear gradient in MPI score change: households that started out poorer at baseline experienced significantly larger reductions in multidimensional poverty by endline, compared to those who were non-poor at baseline.

Rwanda and Uganda show the strongest declines in multidimensional poverty
Changes in MPI scores between baseline and endline varied significantly across countries. Rwanda (-0.23) and Uganda (-0.20) showed the greatest improvements, followed by Ghana (-0.16) and Ivory Coast (-0.07).

Households without partners show greater improvement, illiterate participants less.
Households where the respondent did not have a spouse or partner present showed larger reductions in MPI, suggesting they experienced greater improvements in living conditions compared to those with a partner in the household. Households where the head was not literate showed a smaller improvement in poverty than literate household heads.

Strong improvements in household well-being, especially through increases in home ownership and poverty reduction
Households experienced notable socio-economic improvements, with significant gains in land ownership, homeownership, and reduced debt. We observed a rise in homeownership and a slight increase in literacy among household heads. At the same time, multidimensional poverty dropped significantly: average MPI scores halved (0.38 to 0.21, $p < 0.001$), and the proportion of MPI-poor households was reduced from 69% to 17%. The share of households classified as "non-poor" more than quadrupled. These trends indicate that households not only improved their current living standards but also strengthened their asset base and long-term resilience.

Nutrition is the primary driver of multidimensional poverty reduction, with country-level differences as secondary factors
The largest contributor to MPI improvements among those exiting poverty was nutrition (improvements in HFIAS) , accounting for over half (52.5%) of the measured total poverty reduction across the pooled sample. Other notable contributors included assets, electricity, and school attendance, while sanitation, child mortality, housing, and water played mid-level roles. Years of schooling and cooking fuel contributed little to overall change. Disaggregated country analysis shows that nutrition consistently drives improvements across all countries, though the relative importance of other components varies. For example, asset gains and reductions in child mortality were more prominent in some contexts than others. Cooking fuel remained a minimal contributor everywhere, as most households were poor in this dimension at both baseline and endline.

4 Benchmarking

4.1 Objectives

The benchmarking exercises uses the available data to estimate what monetary effect (consumption) the 100Weeks intervention could have if implemented in a population of Ugandan coffee farmers.

To estimate the potential effects of the 100Weeks program on populations like those targeted by HereWeGrow (HWG), we model predicted changes in consumption using available monitoring and evaluation data. Specifically, we leverage the Raising the Village (RTV) Uganda baseline dataset, collected in 2024, which captures a rural population aligned with HWG's target group and includes harmonized indicators for constructing a Multidimensional Poverty Index (MPI). To enable comparison across datasets, we construct a 'reduced' MPI using only variables common to both the RTV and 100Weeks datasets.

Using the RTV dataset, 1) we model household daily per capita consumption in 2017 PPP USD based on the reduced MPI components, Household Dietary Diversity Score (HDDS), and household size. 2) We then apply these coefficients to the 100Weeks dataset to predict monetary values of consumption. Consumption is used as the primary welfare metric due to its stability and relevance in rural informal economies, where income is often seasonal or underreported.

These estimations will inform the ROI calculations in the next segment by providing us with monetary (absolute and relative) estimates of impact.

4.2 Data and measurement

For the comparison, we will use a) the 100Weeks dataset for Uganda that was also used in the MPI heterogeneity analysis and b) the Raising the Village (RTV)/HWG baseline dataset. The dataset in a) is described at length in the previous section. We provide a brief description of the second dataset below.

4.2.1 The Raising the Village / HWG dataset (Uganda 2024)

The Pathways Out of Poverty (POP) program is implemented by RTV from 2024 to 2026 in Kitagwenda and Rakai districts, Uganda. The program targets ultra-poor households using a graduation approach, with the goal of improving income, reducing poverty, and enhancing quality of life.

The baseline data was collected by Laterite as part of a stepped-wedge cluster-randomized controlled trial designed to rigorously evaluate the causal impact of the program over time. The dataset includes rich information for 1,260 coffee-farmers on household consumption and livelihood activities, coffee production and agricultural productivity, access to agricultural inputs and credit, and ownership of key assets. It also captures multidimensional aspects of welfare such as food diversity, housing quality, education levels, and health indicators. This comprehensive baseline provides a strong foundation for measuring changes in household welfare and the effectiveness of the POP program in supporting coffee-farming communities to transition out of poverty.

Using the variables common to both the RTV and 100Weeks datasets, we construct a reduced, harmonized Multidimensional Poverty Index (MPI) that allows for direct comparison across the two sources. While the RTV dataset includes observed data on household consumption, the 100Weeks dataset does not. To bridge this gap, we use the RTV data to model the relationship between non-monetary MPI components (including HDDS and household size) and monetary consumption, which serves as a proxy for household welfare.

We then apply this model, specifically the estimated coefficients from a generalized linear model to the 100Weeks dataset, where the same MPI components are available, but consumption is not measured. This allows us to predict daily per capita household consumption (in 2017 PPP USD) for 100Weeks participants using their MPI indicators. Modelling consumption in this way enables us to estimate potential welfare gains in monetary terms and project what kind of increase in consumption might be expected if the 100Weeks intervention were implemented among similar populations, such as Ugandan coffee farmers.

4.2.2 Construction of reduced form MPI

To assess how the 100Weeks program might affect similar populations, such as Ugandan coffee farmers, we construct a harmonized MPI using variables common to the 100Weeks and RTV datasets. This allows us to compare poverty levels and link observed consumption data from RTV to the MPI indicators. We then model the relationship between MPI components and consumption in the RTV data and apply this model to 100Weeks to predict household consumption, enabling us to estimate potential welfare gains in monetary terms.

The RTV dataset has many of the variables necessary to construct the MPI used by 100Weeks. However, there are some notable differences:

1. For RTV dataset there is no information on child mortality for the last 5 years.
2. RTV dataset does not contain data on Household Food Insecurity Access Scale (HFIAS) used for MPI construction. However, both datasets contain the Household Dietary Diversity Score (HDDS).
3. Both RTV and 100Weeks have information on schooling of household head. RTV does not have information on years of schooling completed of all adults in household. MPI uses information on all household members.
4. The asset matrix contains 6 items for MPI construction. There are three common items in both datasets, plus some other items that both surveys have that are not in the MPI construction.

We summarize the variable comparability in the Annex (Table 20).

To construct a reduced Multidimensional Poverty Index (MPI) using the RTV dataset, several adjustments were necessary due to gaps in comparability with the 100Weeks dataset and the standard MPI indicators. The RTV dataset lacked key components such as under-five child mortality data and HFIAS-based food insecurity measures. As a result, child mortality was omitted entirely from the reduced MPI, and food insecurity was indirectly modelled using the Household Dietary Diversity Score (HDDS).

To approximate the nutrition dimension in the RTV dataset, we used an ordered logistic regression to model the relationship between dietary diversity (HDDS) and household food security status (HFIAS categories) in the 100Weeks dataset. In simple terms, this approach estimates how likely a household is to fall into each food security category based on its HDDS score. We then applied this model to the RTV dataset to predict food security levels for those households. This allowed us to include a comparable nutrition indicator in the MPI calculation, even though the original HFIAS data was not available in RTV.

For other MPI dimensions, harmonization and substitution strategies were used to ensure consistency. Standard of living indicators such as sanitation, drinking water, electricity, housing materials, and cooking fuel were aligned across datasets and converted into binary deprivations following standard MPI protocols. In the asset dimension, only three MPI-aligned assets (radio, TV, and phone) were common to both datasets, and two additional items (wheelbarrow and machete) were included to broaden coverage. For vehicle ownership, which was not directly observed in RTV, expenditure on vehicles was used as a proxy. Education was harmonized by using data on the household head’s years of schooling in RTV, instead of full household-level education data. Despite these limitations, the constructed reduced MPI captures core aspects of multidimensional poverty and allows for meaningful comparison across datasets.

4.2.3 Validity of reduced MPI using 100Weeks data

In our context, validity means that the harmonized MPI is a reliable way to reflect the poverty levels we want to measure. This is important because we need to be sure that our results are not influenced by how we built the indicator. Our goal is to compare outcomes across datasets, so the poverty measure must be consistent and meaningful in both cases.

To check this, we compare the harmonized MPI to the original MPI in the 100Weeks dataset. If both versions show similar patterns in poverty levels and changes over time, it suggests that the harmonized MPI is capturing the same underlying concept and can be used with confidence in our analysis.

We present the results of the construction of the reduced MPI, in comparison to the complete MPI used by 100 weeks.

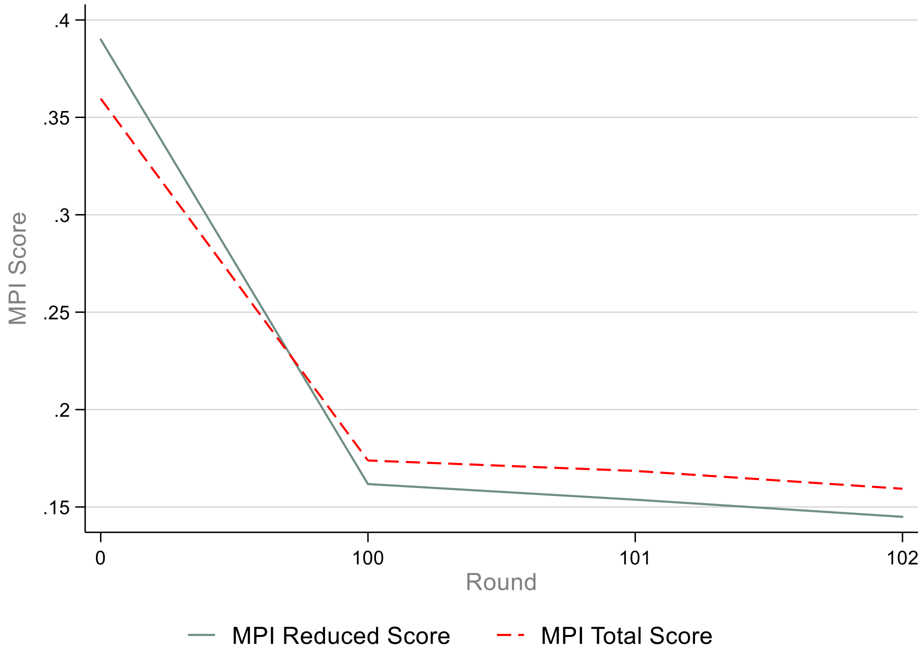


Figure 15 Average MPI score by round – Uganda. Round 0 is BL, round 100 is EL (100 weeks after BL).

Figure 15 shows average MPI score changes by survey round in Uganda, comparing the full MPI and the reduced MPI. At baseline (Round 0), there is a small overestimation by the reduced MPI. The full MPI score starts at approximately **0.36**, while the reduced MPI begins higher, around **0.39,** an overestimation of about **0.03 points**, or roughly **8%** of the full MPI value at baseline. This gap is most pronounced at baseline.

After baseline, however, both scores decline sharply and converge quickly. By Round 100 (ie. after 100weeks), both the full and reduced MPI scores drop to around **0.17**, and from there continue to decline slowly and in parallel. The reduced MPI consistently remains slightly lower, but the difference narrows to roughly **0.01 points** in later rounds. This suggests that while the reduced MPI slightly overestimates poverty at the start, it performs well in tracking changes over time and mirrors the trajectory of the full MPI quite closely.

In Figure 16, we compare the **full MPI scores for 100Weeks participants at baseline (BL) and endline (EL)**. At endline, the full MPI distribution shifts substantially leftward, **reflecting significant improvements in multidimensional poverty.**

In Figure 17, we examine the **reduced MPI scores at baseline for RTV and 100Weeks**. The RTV baseline distribution (solid green line) is flatter and wider, with a peak around 0.6, indicating a more dispersed and higher-poverty profile than 100Weeks baseline (red solid line), which peaks sharply around 0.3. **This suggests that the RTV sample starts off with higher and more varied deprivation levels. In contrast, the 100Weeks endline curve (solid orange line) shows a large shift toward lower MPI scores, clustering around 0.1, illustrating a strong reduction in poverty**. The comparison highlights the initial differences in the populations and the scale of impact observed in the 100Weeks cohort.

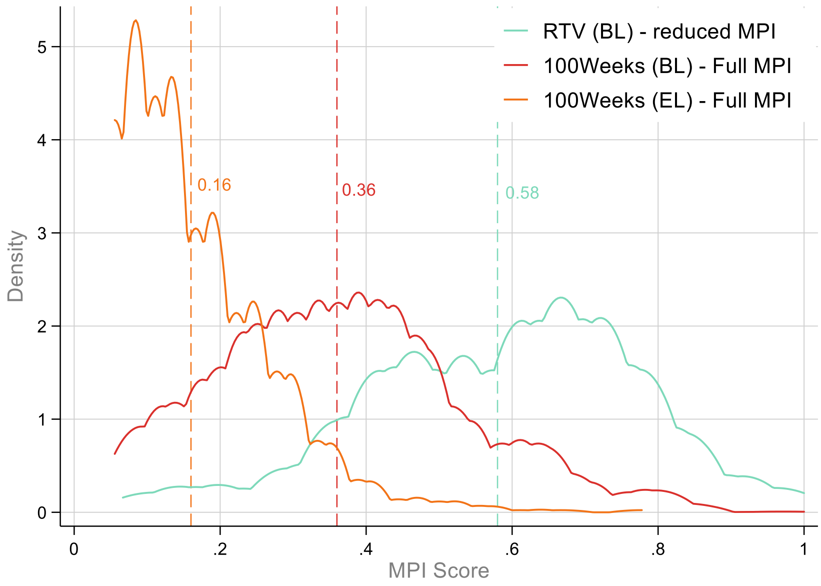


Figure 16 Full MPI for 100 Weeks and reduced MPI for RTV by round. Average MPI score presented in dashed line

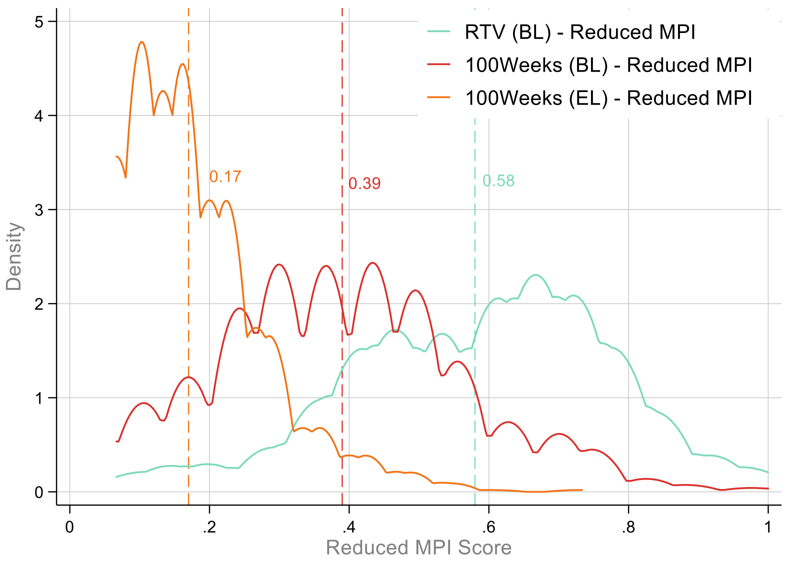


Figure 17 Reduced MPI for 100 Weeks and reduced MPI for RTV by round. Average MPI score in dashed line.

The figures above also provide an idea of the common range across both surveys for reduced MPI scores. Although RTV participants are poorer at BL, there are participants who at each level of reduced-MPI. This is important for relating MPI components to consumption. We explore this in the next section.

4.3 Consumption modelling and prediction

4.3.1 Modelling consumption using MPI components – RTV dataset

In this section, we test whether it is possible to predict household consumption based on MPI components and related indicators. The RTV dataset contains observed consumption data, which allows us to check how well our modelled predictions match the actual values. This step is important because we later apply the same model to the 100Weeks dataset, where consumption data is not available. By testing the model in RTV first, we can assess its accuracy and ensure it is a reasonable tool for predicting consumption elsewhere.

To do this, we model daily per capita consumption using the individual components of the reduced MPI, along with HDDS and household size. Using the individual components rather than the combined MPI score improves prediction accuracy by preserving more detailed information. HDDS is used as a direct measure of nutrition, which is available and comparable across both datasets. Given that consumption data is right-skewed, we use a generalized linear model with a logarithmic link to account for this distribution. The steps of the modelling process are outlined below.

1. We used daily per capita consumption as the dependent variable and HDDS and the rest of the reduced MPI components and number of household members as independent variables, to obtain regression coefficients from GLM modelling using log link function. We excluded fuel from the regression due to lack of variance.
2. The regression coefficients from the RTV dataset were then used to predict consumption in the 100Weeks dataset.
3. We bootstrapped uncertainty in the predictions of 100Weeks.

Table 8 Generalized Linear Model (GLM) coefficients[8]

Predictor[9]	Coef.	z	p-value	Sig.
Household Dietary Diversity Score (HDDS)	0.12	10.06	0.000	***
Head: ≥6 Years of Schooling	-0.03	-0.60	0.551	
School Attendance (All HH Members)	-0.16	-2.28	0.023	**
Access to Electricity	-0.01	-0.17	0.865	
Access to Improved Water Source	-0.08	-1.10	0.270	
Adequate Housing	-0.06	-1.43	0.152	
Asset Ownership Index	-0.12	-2.63	0.008	***
Household Size	-0.10	-5.39	0.000	***

[8] We used daily per capita consumption in 2017 PPP USD as dependent variable using the RTV dataset (n=1260). Significance levels: *** p < 0.01, ** p < 0.05, * p < 0.1.
[9] HDDS is a numeric variable where higher values correspond with values in dietary diversity. Household size is also numeric. The rest of variables are binary, with 1 being coded to represent a negative outcome (poor if value is one). This is why all coefficients are negative for the different components of MPI.

Table 8 presents the results from a generalized linear model (GLM) with a gamma distribution and log link function, using daily per capita consumption in 2017 PPP USD as the dependent variable.

Among all predictors, dietary diversity shows the strongest positive association with consumption. Households with more diverse diets tend to have higher levels of consumption, which aligns with expectations since dietary diversity often reflects greater purchasing power and access to food. On the other hand, larger household size is linked to lower consumption per person, which suggests that resources are more thinly spread in bigger households. Asset ownership also plays a significant role: households with fewer assets tend to have lower consumption levels. Interestingly, households where all members attend school are associated with lower consumption. Other variables such as access to electricity, improved water, and adequate housing show weaker associations and do not appear to be strong predictors of household consumption in this model.

While the coefficients indicate the direction and strength of each association, their actual magnitudes are more difficult to interpret directly because the model uses a log link function. This means the values reflect percentage changes in expected consumption rather than absolute changes. Therefore, we emphasize the relative importance of the predictors and the overall pattern of associations rather than the exact numerical size of the coefficients.

Figure 18 Daily per capita consumption for the RTV BL dataset. Observed and predicted with GLM model. Medians displayed as dashed lines.

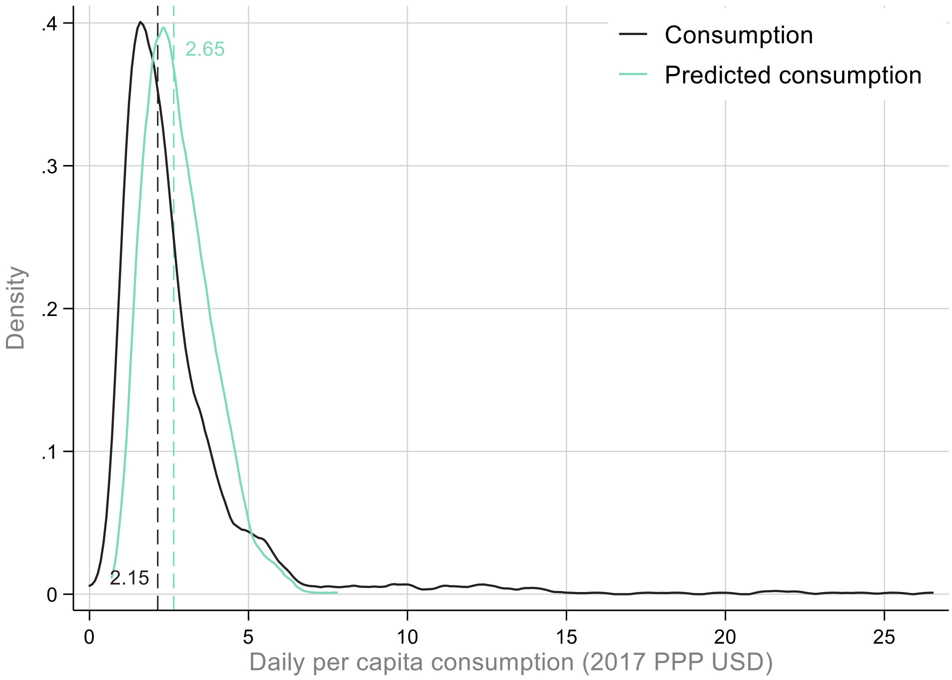


Figure 18 compares the distribution of observed and predicted daily per capita consumption (in 2017 PPP USD) for the RTV dataset. The black line represents the observed consumption data, which is right-skewed with a long tail, reflecting a small number of households with relatively high expenditures. The green line shows the predicted consumption values generated from the GLM model.

Both distributions share a similar shape, with a pronounced peak and rightward skew, indicating that the model captures the general pattern of consumption well. However, the predicted values are more tightly clustered, slightly underestimating the dispersion at the higher end of the distribution — suggesting the model performs well for the bulk of the sample but may underpredict outliers in the upper tail.

The median observed consumption is 2.15 USD, while the median predicted consumption is higher at 2.65 USD (0.50 USD difference **or 24% overestimation**). This difference implies that the model overestimates the central tendency of consumption, even though it aligns closely in terms of overall shape. The medians, plotted as vertical dashed lines, reinforce this observation and highlight the model's solid performance in capturing central values while being slightly conservative on extreme outcomes. The mean values of the two distributions are similar. **This overestimation will be accounted for when using it in the next segment for the ROI.**

4.3.2 Predicting consumption using MPI components – 100Weeks dataset

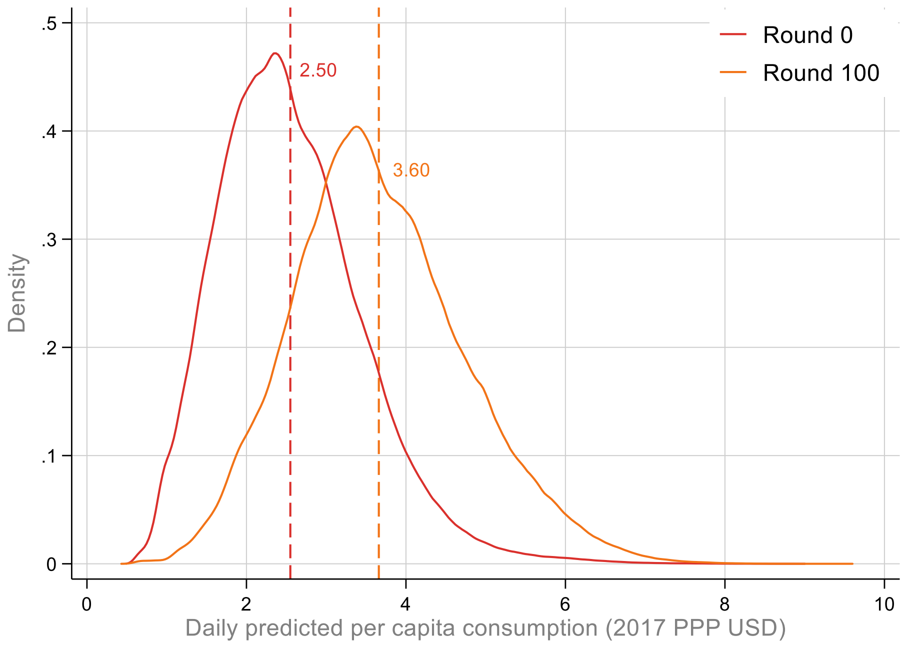
In this step, we estimate what household consumption might have looked like for participants in the 100Weeks program. The 100Weeks dataset does not include direct information on household consumption, so we use information from the RTV dataset to fill this gap. From the RTV data, we learned how household characteristics such as MPI components, dietary diversity, and household size relate to actual consumption. We then use this relationship to predict consumption levels for each household in the 100Weeks data, based on those same characteristics.

To capture the uncertainty around these predictions, we use a method called bootstrapping. This involves repeating the prediction process many times, each time using a slightly different version of the RTV dataset. This helps us understand how much the results might vary depending on the sample. We repeat this process 2000 times, which allows us to generate a range of predicted consumption values for each household. From this, we calculate a confidence interval that shows the likely spread of outcomes.

The results, shown in Figure 19, display the predicted distribution of household consumption in the 100Weeks dataset before and after the program. The main takeaway is that predicted consumption increases noticeably over time, suggesting a positive change in household welfare among program participants. The bootstrapped estimates give us more confidence in this result by showing that the increase holds across many possible scenarios.

For the second step, we used the coefficients derived from the GLM modelling presented in Table 8 to predict the corresponding consumption values in the 100Weeks dataset, using the harmonized components we used to construct the reduced MPI. We bootstrapped the GLM modelling in the RTV dataset 2000 times, and used this to construct a distribution of consumption estimates for the 100Weeks dataset. Figure 19 shows the predicted bootstrapped consumption distributions for the 100Weeks dataset by round. 95% CIs are derived from the 2.5 and 97.5 percentiles of the bootstrapped consumption distribution.

Figure 19 Predicted daily per capita consumption for the 100Weeks dataset by round. Mean values shown in dashed lines. Distributions show bootstrapped estimates using 2000 iterations



The distribution at baseline is centered around a mean of approximately 2.55 (95% CI: 1.13 - 4.46) USD, while the endline distribution shifts noticeably to the right, with a higher mean of about 3.57 (95% CI: 1.81 –5.89) USD. This rightward shift indicates a substantial increase in predicted daily per capita consumption over the program period. While both distributions remain right-skewed, the endline values are more dispersed, reflecting greater variability in predicted outcomes.

The mean increase in predicted daily per capita consumption is estimated at 1.10 (95% CI: 0.89 to 1.33) PPP USD between BL and EL. This amounts to a 44% increase in consumption.

An overview of the baseline consumption values by MPI quintile can be found in the Annex (Figure 26).

Key insights from chapter 4:

Reduced MPI is a reliable proxy for full MPI
The reduced MPI closely mirrors the distribution and changes observed in the full MPI, validating its use for impact analysis when full data is not available.

RTV participants are poorer and more varied at baseline
Compared to 100Weeks participants, the RTV sample starts from higher and more dispersed levels of deprivation, supporting its use as a relevant benchmark for comparison.

HDDS, asset ownership, and household size are strong predictors of consumption
These components of the reduced MPI model are statistically significant and explain part of the variation in predicted daily per capita consumption.

The prediction model performs well on overall distribution
While slightly overestimating the median and underpredicting the upper tail, the model replicates the general shape and pattern of consumption accurately.

Predicted consumption increased significantly from baseline to endline
The mean daily per capita consumption rose by 44% or 1.10 USD (2017 PPP), suggesting substantial economic gains during the program period.

MPI score categories correlate well with predicted consumption
Households with higher MPI scores (poorer) consistently show lower predicted consumption, reinforcing MPI's validity as a proxy for economic status.

Larger poverty reductions link to greater consumption gains
Households that experienced the greatest drops in MPI also saw the largest increases in predicted consumption.

5 Return on investment

5.1 Objectives

The objective of this chapter is to estimate the potential ROI consumption of a 100Weeks program in Uganda. The ROI is calculated as the ratio between program costs and benefits. Benefits are expressed as yearly household consumption gain per beneficiary treated (in 2017 USD PPP). Costs are also expressed per beneficiary treated and in the same monetary unit.

5.2 Overall assumptions

We treat the Multidimensional Poverty Index (MPI) as a household-level indicator. Since we project consumption based on MPI scores, we consider both MPI and predicted consumption as household-level metrics. As such, household consumption is used as the primary welfare outcome in our analysis.

Although the program targets individual women, we assume income gains and resulting investments (e.g. in housing, education, or food security) affect the household. This reflects an underlying theory of change that income diversification by a woman benefits the broader household, not just the individual.

We project benefits and costs over a 10-year period. The consumption gains observed between baseline and endline are assumed to decrease by 3% annually for the main analysis.

This analysis assumes that the observed pre-post consumption change is fully attributable to the program. However, this is not a counterfactual-based estimate and does not account for secular trends or external influences. As such, it likely overestimates the true impact and is not directly comparable to cost-effectiveness estimates in the literature that rely on rigorous causal inference methods.

All costs and benefits are expressed in 2017 PPP USD for consistency with global benchmarks. To convert the costs of the program based on Uganda 2024, we first adjusted euros spend in 2024 to Ugandan shilling (UGX) and then deflated them to 2017 UGX and used purchase power parity (PPP) conversion. This ensures comparability across time and regions in real purchasing power terms.

5.3 Cost assumptions

This section explains how the annual costs per program participant is derived. Costs of 100Weeks are presented following the widely used J-PAL methodology. These costs are then divided by the number of participants and expressed in internationally 2017 USD PPP.

5.3.1 Cost following the JPAL framework

The Abdul Latif Jameel Poverty Action Lab (J-PAL), a research institute founded by researchers from Massachusetts Institute for Technology (MIT), has developed a costing template to guide cost-effectiveness assessments. This template is widely used in impact evaluations and gives a standardized way of comparing costs across different contexts and interventions. The template builds on the 'ingredients method, as postulated by Dhaliwal et al. (2013). Costs are categorized into 8 ingredients, that are all related to a specific project, rather than to running an organization in its entirety. The basic idea is that the costs should cover the **incremental costs** of a project, that would be incurred when scaling or repeating the program. This includes, for example, the costs of program administration, training of staff involved in the intervention, but not the costs of general fundraising or salary of the managers of a head office[10].

The costs data provided by 100Weeks is based on the actual costs made in Uganda in 2024, in euros (2024 conversion rates). For some items, rule of thumb' calculations have been made, in line with the JPAL methodology. For example: the annual salary of the program director in Uganda – included under *1. Program administration and staff costs* – was included for 67%, as this program director was spending around 33% of the time on general fundraising-related tasks for the head office, not related to the program in Uganda. Occasionally, the figures from the bookkeeping are not an accurate reflection of the actual costs made, and figures from the budget are used as a substitute[11].

Table 9 Program costs in Uganda 2024 (J-PAL method)

J-PAL Cost Items:	100Weeks Costs in Uganda 2024
1. Program administration and staff costs Cost of all full-time staff who worked **throughout any phases of the intervention and implementation** and other costs related to program administration.	€ 80,399
2. Targeting costs* Costs that were incurred **to target, identify, and raise awareness among potential subjects as part of the intervention.** Targeting costs may include costs of a pre-program census or targeting survey given to identify those within a specific region who are eligible and meet certain criteria. This category also includes marketing costs, such as the costs incurred to print and distribute flyers or host information sessions.	€ 1,823* (this number includes salary of enumerators in targeting. Lists of potential participants are provided by partner organizations)
3. Staff Training	€ 857

[10] This is the template used (version 3.1.3, April 2017), guided by these guidelines from J-PAL. A new version (version 4.1, April, 2024) has been published by J-PAL.

[11] E.g. for item 8, monitoring costs, the figure from the bookkeeping was too low and is corrected based on their budget figures. This judgement is made by the director Peter Meijer, director of 100Weeks, in coordination with their accountant. 100Weeks has shared an excel-excerpt from their bookkeeping of Uganda 2024 with Laterite. For further clarification or details on specific cost items, we recommend contacting 100Weeks.

J-PAL Cost Items:	100Weeks Costs in Uganda 2024
Costs that were incurred to train staff involved in the intervention. This does not include training for enumerators who conducted surveys to collect data for program evaluation.	
4. Participant Training Costs incurred by the program implementer to train participants or beneficiaries.	€ 25,118
5. Implementation, cash transfer and program material costs Costs of implementing the intervention. This can include the cost of items distributed to participants, the cost of distributing the items, staff transportation to provide services/implement the program, or the cost of creating and maintaining technologies or resources developed for the intervention.	€ 596,051
6. User costs Costs that the user incurred as a part of the intervention. This tab also includes the opportunity cost of participants' time, so interventions requiring a large time commitment from participants should fill out this tab, even if there are no other user costs.	2 hours every week for training, no monetary value assigned
7. Averted costs Costs averted as a result of the intervention. Only include costs here that are significant.	n/a
8. Monitoring costs* Costs incurred to oversee and monitor program activities, or track program recipients or staff and their progress during the intervention. This category also includes costs of monitoring supply chains or other systems set up for the intervention. Please do not include costs for data collection for monitoring or evaluation which would not take place in a full-scale version of the program.	€ 30,917*
Total program cost The total program costs include costs across various categories, such as Program Administration, Targeting, Staff Training, User Training, Implementation, User Costs, Averted Costs, and Monitoring.	**€ 735,165**

**Targeting and monitoring costs are added under the same ledger in the bookkeeping of 100Weeks in Uganda in 2024 and are disentangled based on estimations. The targeting costs per starting participant are multiplied by the number of participant-years to get a representative cost (see below for an explanation of participant-years), and divided by 1.92 (a full program of 100 weeks are 1.92 years) to even out the targeting costs over the full duration of the program, as these same numbers are also used as costs in program year 2.*

5.3.2 Converting the annual costs

To obtain accurate costs per program participant, we need to take into account that the 100weeks program lasts slightly shorter than two years (100 weeks instead of 104). Therefore, we divide the annual total costs in Uganda over the number of participant-years that were in the program in 2024. A participant-year equals a beneficiary that is in the program for the entire year. Beneficiaries participating for only a portion of 2024 are counted proportionally. For example: two participants that were both in the program for half a year, count as one participant-year together. This calculation, performed by 100Weeks, was based on the number of transfers made in Uganda in 2024. The total sum of transfers is divided by 416 (52 weeks times €8) and gives 1438 participant-years[12]. The total cost of **735,165 EUR divided by 1438, gives a cost of 511 EUR (2024 EUR) per participant-year.**

The costs of 2024 in Uganda of 511 EUR equals 2,078,656 UGX based on the exchange rate of 2024[13]. This is converted to January 2017 prices, based on the inflation data provided by the IMF[14]. 2,078,656 UGX in 2024 equals 1,533,873 UGX in 2017 price levels, as prices are 36% higher in 2024 compared to 2017. This 1,533,873 UGX is divided by 2019, the PPP-conversion factor of 2017 provided by the United Nations[15], giving an **annual cost per participant of 1258 USD in 2017-PPP-USD.**

5.4 Benefit assumptions

Objective: get a benefit estimation that allows for comparison across space and time, following a standardized, internationally used method. Description of the general process

For the first and second year, benefits are limited to the value of the cash transfer received by participants. This is because available data only include consumption measurements at baseline and after two years (i.e., post-program completion under the 100Weeks model), not after one year. To estimate the impact of the program on consumption, we calculate relative gains: a 41.6% increase in predicted **median (1.04 increase in 2017 PPP USD consumption)** of daily per capita consumption between baseline and endline. We use the median to as a more conservative metric, but the difference between mean and median is not substantial (1.10 USD mean increase, versus 1.04 for the median). We decided to use the relative amount to capture the increase, which is then applied to a corrected baseline value of consumption (due to overestimation at the modelling of consumption).

We use a baseline consumption of 2.50 USD PPP as per the estimates of mean consumption in the benchmarking segment. Since predicted consumption slightly overestimates actual consumption in the RTV dataset (by approximately 0.50 USD in 2017 PPP, or 25%), we correct for this at the baseline consumption part (2.50 * 0.75= 1,88). The final benefit is expressed as a **yearly increase in daily per capita consumption of 0.78 USD PPP** that amounts to 41.6% the baseline consumption of 1.88 PPP USD. The yearly benefit starts to accrue at the beginning of year 3, and decays at a rate of 3%. We convert to household income by multiplying by the average household size in the 100Weeks dataset (0.78 * 4.3 members

[12] The sum of transfers made (€598,208) is very similar to the costs listed under *5. Implementation and program material costs* (€596,051). This difference of 0.36% is negligible, but unexplained. Our double-check on participants for which these costs in Uganda in 2024 are incurred, using the date of the baseline interview, yielded very similar figures for program participants.

[13] Using the average exchange EUR/UGX exchange rate of 2024, 1 EUR equals 4065.91 UGX.

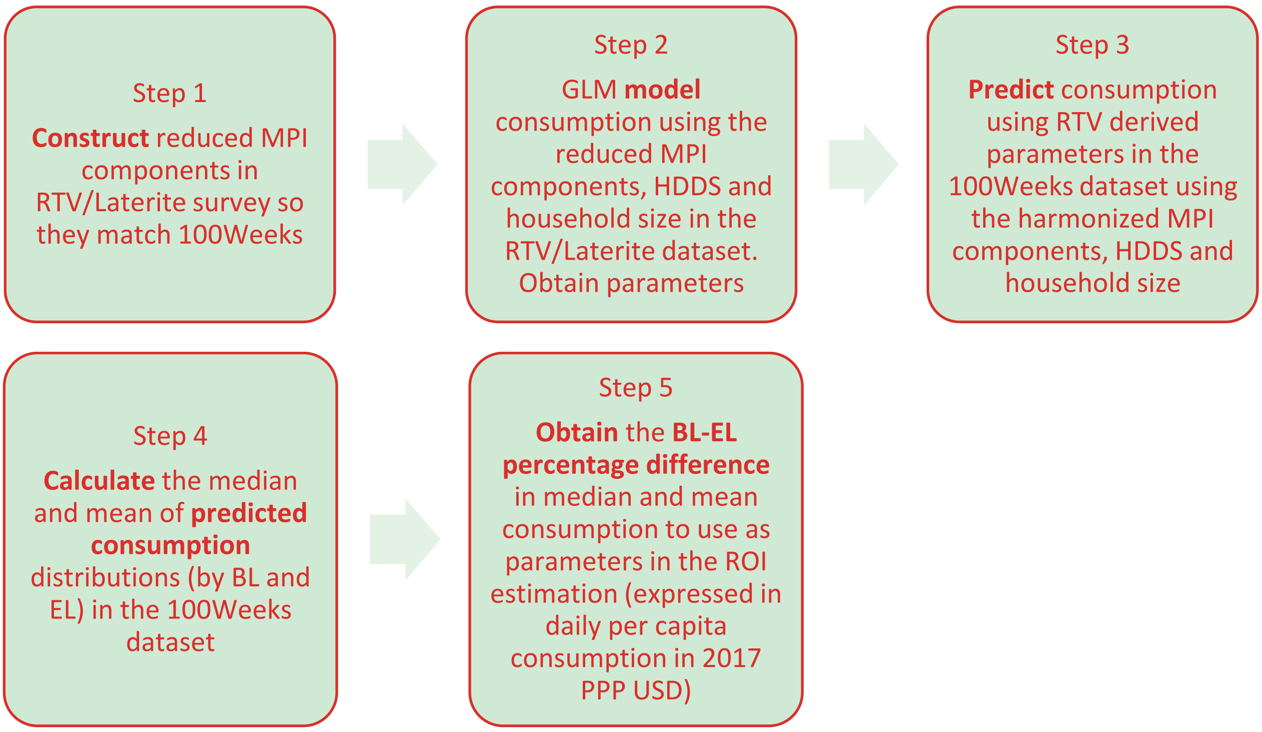
[14] The inflation in Uganda averages at around 3.9% per year over 2017-2014 (https://www.imf.org/external/datamapper/PCPIPCH@WEO/UGA).

[15] See the Ugandan PPP conversion factor for 2017 that converts Local Currency Units (LSU) to USD-PPP, based on private consumption, here: https://data.un.org/Data.aspx?d=WDI&f=Indicator_Code%3APA.NUS.PRVT.PP

= 3.35 USD per household per day)[16]. Finally, we multiply by 365 days to convert it to yearly benefit (3.35 * 365 = **1224 2017 PPP USD per household per year**). We also present a range of estimates, based only on variation of the yearly benefit as per the benchmarking segment (the 95% CI of the difference in consumption), to partially express the uncertainty in the estimates. A graphic representation of the process is presented in Figure 20 and Figure 21 below.

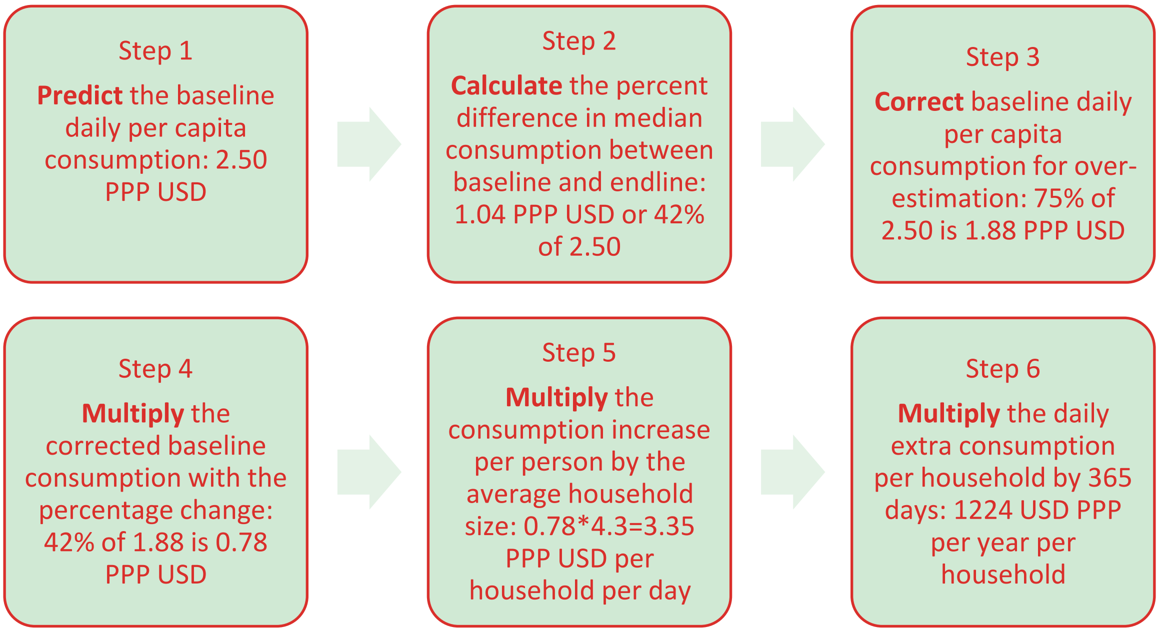
An RCT of a cash transfer programme implemented by GiveDirectly in Uganda that disbursed an amount of $1000 USD to Ugandan coffee- farmers over the course of four months[17] estimated an increase of monthly consumption of 40%, which translates to 99 USD monthly in 2011 PPP USD (112 PPP 2017 USD). This is close to our estimate of 101 PPP 2017 USD found using our modelling, both in relative and absolute terms. Although these estimates are not directly comparable as they are based on an RCT and incorporating a counterfactual group, the consumption increases estimated are of a similar order of magnitude.

Figure 20 Process to derive BL-EL consumption change



[16] A more conservative approach could be to use adult equivalent instead of household size. For the 100Weeks sample average AE was 2.8 individuals. This translates to 2.34 USD per household per day, and 803 USD per year. This translates to a lower ROI (2.24) than our main estimate (2.99).
[17] Michael Cooke and Piali Mukhopadhyay (2019). *Evaluating large cash transfers to coffee farmers: Evidence from eastern Uganda*, GiveDirectly and IDinsight.

Figure 21 Process to derive benefits per year per household from per capita daily consumption



To understand the ROI in terms of the variation in change of MPI that the heterogeneity analysis showed, we then propose a variation of results, using the MPI quintile change between BL and EL (1 is highest decrease, and 5 is smallest decrease or even increase in MPI). We then present ROI based on the corrected baseline consumption for each of these quintiles, as well as the estimated relative change between BL and EL. This shows the heterogeneity in results that is concealed in the average calculation. We use the following parameters for each quintile, using a similar procedure of conversion as described for the mean scenario.

Table 10 MPI quantile change and its corresponding increase in consumption

MPI change quintile	% increase in consumption between BL and EL	BL per capita daily consumption (corrected) in 2017 PPP USD	Absolute increase in consumption in 2017 PPP USD	HH yearly increase in consumption in 2017 PPP USD
1 (largest)	0.77	1.66	1.28	**2008.89**
2	0.42	1.98	0.83	**1297.93**
3	0.31	1.94	0.60	**935.93**
4	0.36	2.01	0.73	**1141.07**
5 (smallest)	0.25	2.06	0.51	**798.81**

5.5 Scenarios

5.5.1 ROI calculations using mean value of consumption change

Table 11 ROI estimation based using means of parameters

	Costs		Benefits	
Year	Step 1: costs per year	Step 2: Discounted (10%)	Step 1: Decayed Benefits (3% annually)	Step 2: Discounted Benefits (10%)
1	1258.11	1258.12	1023.74	1023.74
2	1161.23	1037.95	944.99	850.49
3	0	0	1187.28	961.70
4	0	0	1151.66	839.56
5	0	0	1117.11	732.94
6	0	0	1083.60	639.85
7	0	0	1051.09	558.59
8	0	0	1019.56	487.65
9	0	0	988.97	425.72
10	0	0	959.30	371.65

Present Value of Costs: 2303.11

Present Value of Benefits: 6891.90

ROI: 2.99 (2.57 – 3.59)

Table 11 above presents a 10-year Return on Investment (ROI) analysis, incorporating both the decay of program effects over time and the discounting of future values. For the first two years, the benefit is assumed to be the transfer value. From year 3 onwards, we apply an annual effect decay of 3% (decay factor = 0.97) to the annual modelled benefit of 1224 PPP-adjusted units, while costs are assumed to occur in the first two years only. This decay reflects the assumption that the impact of the program diminishes slightly each year. In Step 2, we apply a discount rate of 0.9 (10%) to account for the time value of money, adjusting both benefits and costs accordingly. The result shows that the present value of total program costs is approximately 23013.11, while the present value of benefits reaches 6891.9 in 2017 PPP USD. This yields an estimated **ROI of 2.99,** meaning that for every unit of cost incurred, the program is expected to generate nearly three units of social benefit over the 10-year period. If we use the upper 95%CI of the consumption difference prediction, we have the boundary presented in parentheses ranging from 2.57 to 3.59 assuming the largest change. This does not represent a 95% CI for the ROI estimate since it incorporates only the uncertainty of the consumption prediction on ROI estimation by MPI change quintile.

We conducted similar ROI calculations by MPI change quintile, but show results only for the estimated ROIs.

Table 12 ROI by MPI quintile change between BL and EL

MPI change quintile	% increase in consumption between BL and EL	BL daily per capita consumption (corrected for overestimation at modelling stage) in 2017 PPP USD	Absolute increase (daily 2017 PPP USD per person)	HH yearly increase in consumption (2017 PPP USD)	ROI
1 (largest)	0.77	1.66	1.28	**2008.89**	**4.39**
2	0.42	1.98	0.83	**1297.93**	**3.12**
3	0.31	1.94	0.60	**935.93**	**2.48**
4	0.36	2.01	0.73	**1141.07**	**2.84**
5 (smallest)	0.25	2.06	0.51	**798.81**	**2.23**

Table 12 above shows the quintile of changes in Multidimensional Poverty Index (MPI) and household consumption growth across quintiles of MPI improvement. Households in the top quintile of MPI reduction (quintile 1) experienced the largest gains, with a 77% increase in consumption between baseline (BL) and endline (EL), translating to an absolute annual increase of 2008.89 PPP and a Return on Investment (ROI) of 4.39. In contrast, households in the lowest quintile of MPI improvement (quintile 5) saw only a 25% increase in consumption, corresponding with a yearly gain of 798.81 PPP and a lower ROI of 2.23. A gradient emerges that greater improvements in MPI are associated with higher absolute consumption gains and stronger returns. This suggests that the program is especially impactful for the most deprived households, reinforcing the value of targeting interventions toward those facing the highest levels of multidimensional poverty.

The estimated ROI figures should be interpreted with caution, as they reflect a range of potential benefits with considerable uncertainty. These estimates represent our best guess of what the return on investment could be under certain assumptions, but they are not definitive. Importantly, the analysis lacks counterfactual. There is no group comparison to show what would have happened in the absence of the intervention. As a result, we cannot confidently attribute the observed changes in consumption or MPI directly to the program. Without a counterfactual, the ROI is ultimately speculative, and while informative, it should not be taken as a precise measure of program impact.

5.6 Sensitivity

5.6.1 ROI calculations using mean value of consumption change

Table 13 Sensitivity of ROI to changes in discount rate

Discount Rate	ROI (Benefit-Cost Ratio)
5%	3.59
10% (main)	2.99
15%	2.53

Table 13 presents the sensitivity of the return on investment (ROI) to varying discount rates. At a 5% discount rate, the ROI is estimated at 3.59, indicating a strong return relative to costs. As the discount rate increases, the ROI declines, reaching 2.99 at the main rate of 10%, and further dropping to 2.53 at a 15% discount rate. This pattern reflects the typical effect of discounting future benefits more heavily.

5.6.2 ROI estimation by MPI change quintile

Table 14 shows the sensitivity of ROI estimates to changes in the discount rate, disaggregated by quintiles of MPI improvement between baseline and endline. ROI is highest for households in the first quintile, those with the largest MPI improvements, reaching 5.37 at a 5% discount rate and remaining strong at 4.39 and 3.62 under 10% and 15% discount rates, respectively. ROI declines progressively across quintiles, with the fifth quintile (smallest MPI improvement) showing the lowest returns: 2.62 (5%), 2.23 (10%), and 1.93 (15%).

Table 14 Sensitivity of ROI estimates to changes in discount rate by MPI change quintile between BL and EL (assuming 3% effect decay)

Discount Rate	1 (largest change)	2	3	4	5 (smallest change)	Average
5%	5.37	3.75	2.93	3.40	2.62	3.59
10% (main)	4.39	3.12	2.48	2.84	2.23	2.99
15%	3.62	2.63	2.12	2.41	1.93	2.53

5.6.3 ROI estimation by MPI change quintile, with different decay assumptions

Another major source of uncertainty is the decay of the 'program effect' over time. In our main analysis so far, we assumed an annual effect decay of 3%. In this section, we present the ROI's by MPI change quantile with 2 different assumptions about the decay of the effect: an annual effect *increase* of 3% and an annual effect decay of 15%. This decay of 15% could be interpreted as a stronger decay of the consumption effect every next year as some new alternative income streams decrease or fail. Alternatively, it could be interpreted as a combined effect of (1) a mild decay in the consumption effect; and (2) the control-group catching up to the treatment group, effectively reducing the *extra* consumption of the program participants. It's important to stress that we lack counterfactual data. However, setting a higher effect decay rate can be viewed as partially incorporating the impact of a counterfactual group that would also experience an increase in consumption over time.

Figure 22 below illustrates the growth trajectory of program benefits with 3 different assumptions about the decay of the effect, prior to discounting. The three assumptions are an annual increase in the consumption effect of 3%, an annual decay of 3%, and an annual decay of 15%. Figure 23 takes that one step further by discounting these future benefits with 5%, 10% and 15%, giving a family of functions, with the discounted functions presented by dashed lines.

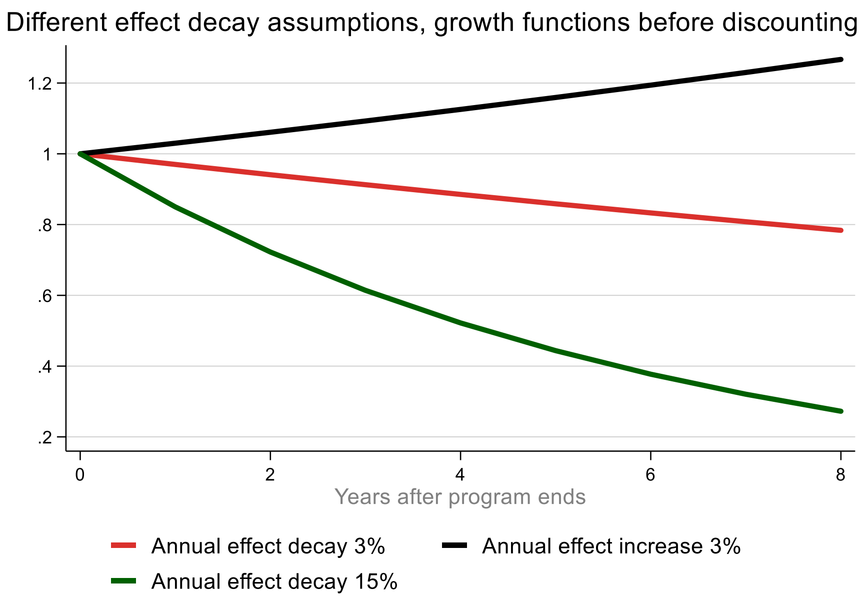


Figure 22 Potential growth functions of benefits, without discounting

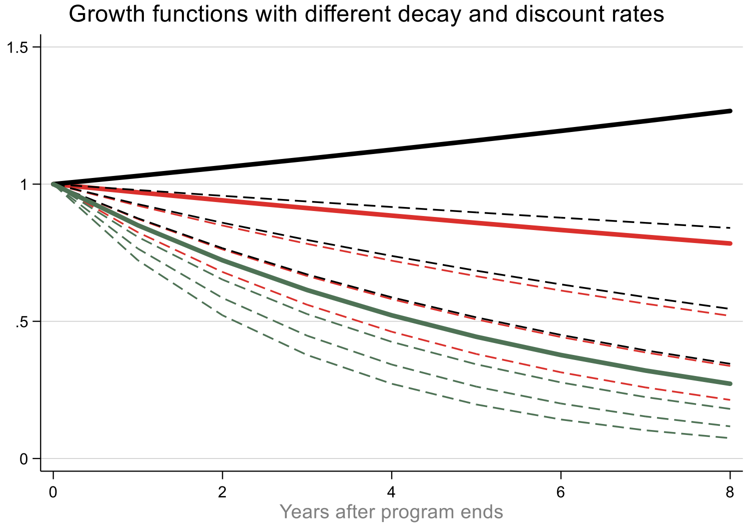
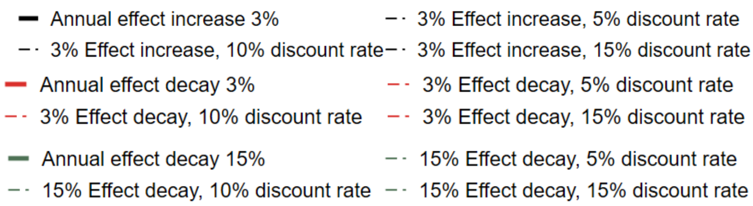


Figure 23 Potential growth functions of benefits, discounted



The decay rate of 3% (red line in figure above) with 5%, 10% and 15% discount rate was already presented in Table 14. Applying the other two decay rates – plus 3% and minus 15% – gives us the following ROI tables below, presenting the different discount rates (5%, 10% and 15%) as well. The average ROIs for a 3% effect increase rate are 4.39 (5% discount), 3.58 (10% discount) and 2.95 (15% discount). The average ROI for a 15% effect decay with 5%, 10% and 15% discount rates are 2.51, 2.19 and 1.93.

Table 15 Annual effect increase of 3%, by MPI change quintile between BL and EL

Discount Rate	1 (largest change)	2	3	4	5 (smallest change)	Average
5%	6.68	4.61	3.55	4.15	3.15	4.39
10% (main)	5.35	3.74	2.93	3.39	2.62	3.58
15%	4.32	3.08	2.45	2.80	2.21	2.95

Table 16 Annual effect decay of 15%, by MPI change quintile between BL and EL

Discount Rate	1 (largest change)	2	3	4	5 (smallest change)	Average
5%	3.59	2.61	2.11	2.39	1.92	2.51
10% (main)	3.07	2.27	1.86	2.09	1.71	2.19
15%	2.65	2.00	1.67	1.86	1.54	1.93

5.6.4 Sensitivity summary

Our baseline analysis estimates a Return on Investment (ROI) of 2.99. This calculation assumes the mean predicted change in consumption, a 10% discount rate, and a 3% annual decay in the program's effect. This sensitivity analysis, which varied the discount rate (5%, 10%, 15%) and the effect decay (+3%, -15%), produced a wide range of ROIs, ranging from 1.93 to 4.39. Importantly, the **ROI remained firmly above 1 across all scenarios, indicating that the program's benefits consistently exceed its costs.**

5.7 Limitations

A key limitation of this analysis is the absence of a counterfactual. Without a comparison group, we cannot distinguish between changes that were caused by the program and those that may have occurred anyway due to other factors such as broader economic trends or seasonal variation. As a result, the analysis assumes all observed improvements in MPI and consumption are attributable to the program, which likely leads to an overstatement of its true impact.

Secondly, the analysis depends on modelling household consumption from MPI components, HDDS and household size, using a conversion derived from a different context (Rakai and Kitagwenda districts of Uganda). While we adjusted this model to better reflect the hypothetical 100Weeks population, it remains an imperfect proxy for actual consumption. The risk of overestimation is especially relevant here, since the RTV sample is poorer than the 100Weeks population.

Finally, we had to make assumptions about how benefits accumulate over time. We have assumed a 3% decay in benefit starting in year three, after the last cash transfer year. In reality, benefits may emerge earlier or fade away in a few years, depending on household behavior and external conditions. The scarce academic literature on long-term (5+ years after the transfers stop) monetary effects of cash transfers in developing countries does not provide a solid basis to assume a certain level of decay of the effect of the 100Weeks program over 10 years. Counterfactual-based research about cash transfers in Uganda (Blattman et. al., 2019[18]) shows that the effect on earnings is still visible at the 5 year follow-up, but dissipates at the 9-year follow-up. Interestingly, at the 12-year follow-up during the COVID-19 crisis, the effect re-emerged (Fiala et. al., 2022)[19]. Not all studies confirm long-lasting income effects. For example, Baird et. al. (2019)[20], show dissipating income effects in Malawi, already 2 year after the transfers have stopped. This rough approach to timing introduces uncertainty in how we calculate both the present value of benefits and the ROI, and highlights the need for more detailed, time-sensitive impact data in future evaluations.

5.8 Recommendations

To improve the accuracy of impact estimates, it would be highly desirable to collect monetary consumption data alongside the MPI already measured by 100Weeks. Given the constraints of time and respondent burden, the use of reduced consumption modules represents the best compromise between data quality and efficiency, particularly in low-income settings where income is often underreported or unstable. If full consumption measurement is not feasible, the survey should at minimum capture data on key income sources such as agriculture, remittances, formal and informal labor, business activities, and livestock-related income (both sales and products). This would allow for better triangulation of economic outcomes and strengthen the credibility of ROI estimates.

The analysis of heterogeneity and ROI clearly indicates that poorer households experience the largest gains from the program, both in terms of absolute consumption increases and returns on investment. Given that the RTV population appears to be relatively poorer than the original 100Weeks sample—yet still shows positive impacts—there is strong justification for intensifying outreach to the poorest segments. Enhancing targeting mechanisms to better reach these groups could significantly improve the overall impact and cost-effectiveness of the program.

[18] Blattman, Christopher, Nathan Fiala, and Sebastian Martinez. 2020. "The Long-Term Impacts of Grants on Poverty: Nine-Year Evidence from Uganda's Youth Opportunities Program." *American Economic Review: Insights 2* (3): 287–304. DOI: 10.1257/aeri.20190224

[19] Fiala, N., Rose, J., Aryemo, F., & Peters, J. (2022). The (very) long-run impacts of cash grants during a crisis (No. 961). *Ruhr Economic Papers*.

[20] Baird, S., McIntosh, C., & Özler, B. (2019). When the money runs out: Do cash transfers have sustained effects on human capital accumulation?. *Journal of Development Economics, 140*, 169-185.

Key insights from chapter 5

The 100Weeks program delivers strong financial returns under multiple scenarios
The estimated ROI is 2.99 under baseline assumptions (10% discount rate, 3% effect decay), with all tested scenarios yielding ROI > 1, confirming that benefits consistently exceed costs.

Program impacts vary significantly by poverty level at baseline

Households with the greatest reductions in MPI (poorest at baseline) achieved the highest ROI (up to 4.39), while households with less MPI improvement had lower returns (down to 2.23). This suggests stronger returns when targeting the most deprived.

Sensitivity analysis confirms robust results

ROI estimates remained strong (1.93 to 4.39) across a range of decay assumptions (from -15% to +3%) and discount rates (5%, 10%, 15%), reinforcing the robustness of the findings.

Long-term effects of cash transfers remain uncertain

The assumed 3% decay in program impact is based on limited literature. Some studies show persistent effects (e.g. Uganda), while others show faster dissipation (e.g. Malawi). More follow-up data is needed.

Data collection should include monetary consumption or income
For stronger estimates, future evaluations should include either reduced consumption modules or detailed income sources, especially in low-income contexts where income is volatile.

6 Conclusion

This report presents a comprehensive assessment of the 100Weeks program, combining internal monitoring data with external evidence and analytical projections. The findings affirm that the program makes a significant contribution to poverty reduction and financial empowerment among women in target communities. However, the strength of the conclusions is hampered by the absence of a credible counterfactual, limiting the ability to attribute observed changes solely to the intervention.

The multidimensional poverty reductions observed in internal monitoring are substantial, with MPI scores and poverty headcounts declining across all countries. These results are promising and indicate that the program reaches its intended beneficiaries and addresses key deprivations such as nutrition, asset ownership, and school attendance. Nevertheless, comparisons with external studies suggest that internal estimates may overstate true impact, while external studies likely underestimate it due to methodological constraints.

The heterogeneity analysis underscores the importance of targeting: households that are poorer at baseline consistently benefit the most, and meaningful variation exists by country, household structure, education, and livelihood type. These insights support a more differentiated approach to programming and suggest the need to tailor interventions more closely to context and participant profile.

The ROI analysis shows that the benefits of the intervention outweigh the costs under a wide range of plausible scenarios. Although assumptions about long-term gains and decay introduce uncertainty, the analysis provides a credible bound and indicates that continued investment in the 100Weeks model is likely to generate meaningful returns in poverty reduction.

Going forward, we identify two core recommendations to enhance both impact and evidence:

1. Improve data systems by adding monetary consumption or income indicators alongside MPI and other multidimensional measures through the inclusion of rapid consumption measures. The implementation of a rapid consumption module is outlined by the World Bank. There is however, no one-size fits all module, since the methodology relies on using national (or regional) public data to identify the most consumed items and form core and alternative modules. This means that this will vary between countries, but also within regions inside countries. For Uganda, an example would be using the National Panel Survey, which contains a full consumption module. The implementation should aim to take the complexity to the analyst and make it as simple as possible for the enumerator.

2. Strengthen targeting efforts to prioritize the poorest, who consistently show the largest gains and the highest return on investment.

7 References

Anne, J. M. (n.d.). *A pathway out of extreme poverty* [Master's thesis]. University of Amsterdam.

The Cash Lab. (2022*). Studying the effects of multifaceted cash transfers on the living-income gap*. KIT; IDH; 100WEEKS; Unilever; Cargill.

Hijmissen-Bonhof, W. H. (2020). *Financial shocks and mental health: An econometric panel data analysis* [Master's thesis]. Vrije Universiteit Amsterdam.

Kell, B. (2020). *Exploring the impact of an unconditional cash transfer programme on cognitive and relational well-being* [Master's thesis]. University of Amsterdam.

Klunder, L. (2020). *There is more to it than cash: The importance of a multi-faceted cash transfer program for women's empowerment, an empirical assessment* [Unpublished master's thesis].

Langener, A. (2020). *The economic impact of unconditional cash transfers in developing countries: Evidence from the 100weeks program in Rwanda* [Master's thesis]. Vrije Universiteit Amsterdam.

Langener, A. (n.d.). *The relationship between income, social support, mental wellbeing and empowerment in developing countries: Evidence from the 100weeks program in Rwanda* [Master's thesis]. University of Amsterdam.

Saleh, M. (2024). *Empowering change: Revealing the impact of cash transfer programs on women's empowerment* [Master's thesis]. Vrije Universiteit Amsterdam.

Wolf, S., Kembou, S., & Ogan, A. (2024). Cash transfers improve economic conditions and reduce maternal stress in rural Côte d'Ivoire. *Journal of Child and Family Studies, 33*(4), 1251–1265. https://doi.org/10.1007/s10826-024-02800-z

ANNEXES

Annexes

Table 17: MPI change for 100W participants and their relevant regional level comparisons

Country	2-year MPI (AF-method) change for 100W participants	Average 2-year change: regional level comparison*	Regional comparison: % of observed pre-post for program participants
Rwanda	-0.3539	-0.0226	6.4%
Ghana	-0.2374	-0.0231	9.7%
Uganda**	-0.2490	-0.0276	11.1%

**The comparison is the average 2-year change within the relevant poor provinces. The provinces are selected following the procedure described above. The provinces are West and South in Rwanda (2 time periods each), Eastern and Northern in Uganda (one period each) and Northern and Upper East in Ghana (two periods each).*
***Data limitations give a less informative comparison point for Uganda, than for Rwanda and Ghana. See explanation below.*

Table 18: Aspects covered regarding data quality of 100Weeks

Topic	Sub-questions
Toc alignment of indicators	To what extent do the indicators within the monitoring framework, and the instruments used reflect the outcomes and pathways posited in the ToC?
Construct validity & credibility of indicators	Additionally, broader concepts —such as food security— can be measured in diverse ways. Does 100Weeks use internationally validated and cross-nationally comparable indicators?
Data quality	How are data collectors selected, trained, and supervised? Does 100Weeks use survey metadata —such as total survey time, time spent per section, and location— to monitor data quality? Are there any measures in place to assess the reliability of responses, such as using different formats —triangulation— for the same question?
Attrition & completeness:	If data is missing in a non-random way, the representativeness of results is compromised (e.g. the poorest people were more likely to refuse). Missing value patterns, and patterns in attrition and (suspected) reasons why data is missing, helps to mitigate this.
MEL-practices	Does the monitoring data inform on project progress? Does it yield actionable insights?

Topic	Sub-questions
Data cleaning	Is the methodology for data cleaning clearly documented?

Transparency and documentation of data collection, data cleaning and calculation of indicator values are integrated in the other sections described in this table. Sampling is not part of the procedure of 100Weeks as they monitor all program participants.

Research documents' contributions to a ToC

All research documents collectively inform our understanding of program effects and causal pathways. Figure 24 visually represents the contributions to a Theory of Change (ToC), showing the scope of the research documents. These contributions can serve as valuable input for developing the causal pathways to program outcomes in more detail.

100Weeks program activities and outputs are summarized in light and dark grey boxes, leading to financial (intermediary) outcomes in blue. Several effects emerge from these financial outcomes and are presented as arrows:

- Black: subsequent step follows by definition;
- Green: positive, intended program effect;
- Red: negative, unintended program effect;
- Dashed grey: effect studied but no significant program effect is detected.

Mechanisms driving these outcomes are shown in boxes along the corresponding arrows. These mechanisms are inferred from both quantified intermediary outcomes and qualitative evidence. References to the relevant research documents appear in brackets; for example, (3) refers to document 3 in Table 1. Final outcomes are illustrated in orange boxes.

Special attention can be paid to the blue circle in the middle with financial intermediary outcomes, considering HWG's focus on income improvement. Disposable income – money available to be spend or saved – is increasing for program participants. This is true by definition during the 100 weeks of the program, as the weekly cash transfer represents extra income which they can allocate as they please. During and *after* the program, the increased income from farming, as a result of investments or new alternative income streams, represents this increased disposable income as well.

A partial knowledge gap is observed in the causal pathways that link the program to changes in MPI. Note that dietary diversity, mental well-being, and the intra-household power dynamics are not part of the MPI. Educational aspirations inform the causal pathway to school attendance, and the path to decreased food insecurity through increased expenditures on basic needs. That leaves many of the 10 sub-components of the MPI still out of scope of the nine external research documents.

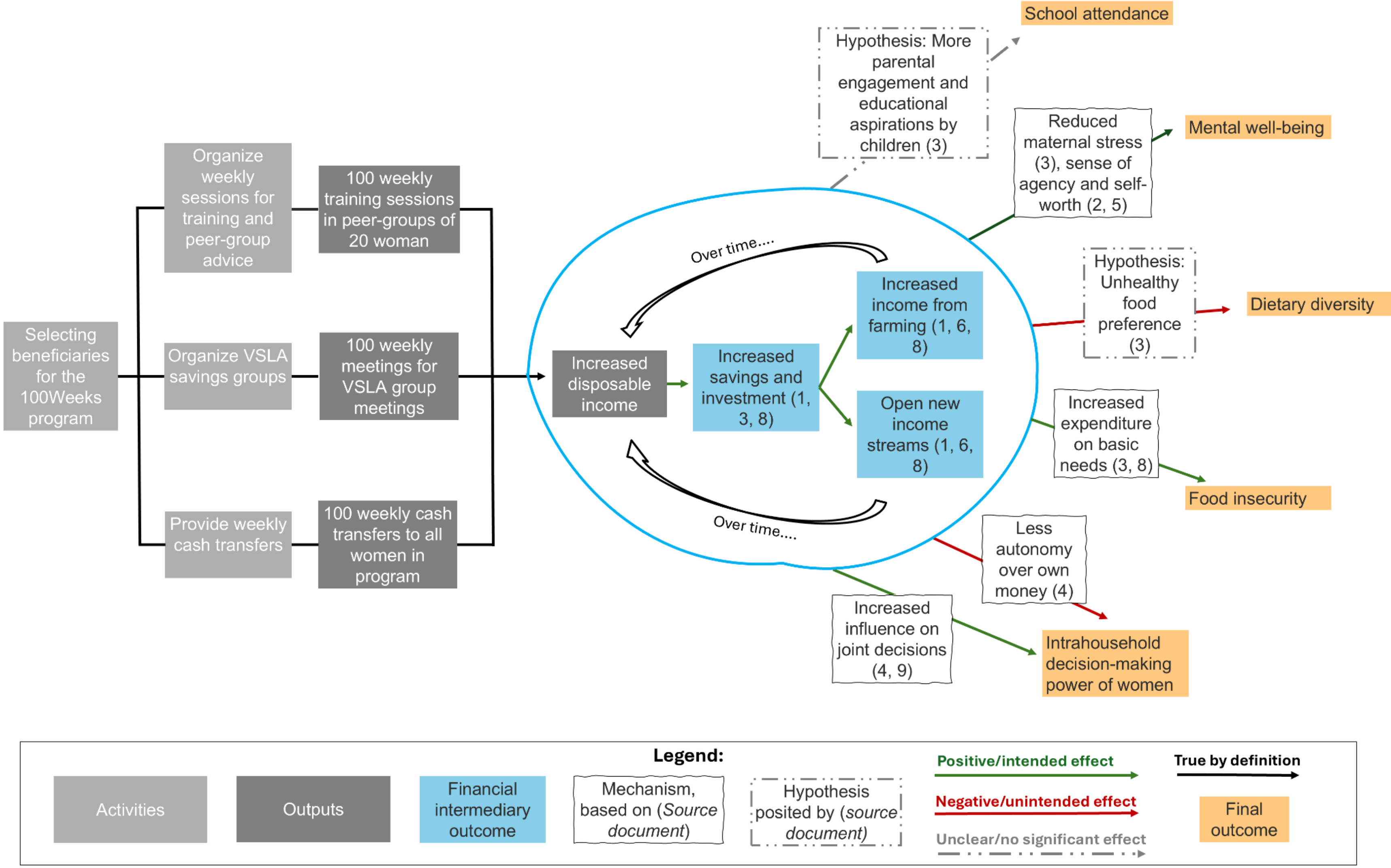


Figure 24 Research documents' (1-9) contributions to a ToC. Source: interpretations of research documents 1-9 by Laterite

Table 19 Scoring research documents

	A. Clarity and articulation of ToC (1-4):		B. Assumptions and causal pathways (1-4):				C. Adaptation and learning (1-4):	D. Rigor (1-4):			E. Rigor: Maryland scale (1-5):
	Does the research report describe a ToC relevant for the 100Weeks model?	Are the core intervention components (cash transfers, training, savings groups) and the hypothesized causal pathways explicitly identified?	Does it explicitly link its measured outcomes to the ToC, addressing the hypothesized causal pathways?	To what extent does the study examine intermediate outcomes or mechanisms predicted by the ToC?	Does the report test the underlying assumptions of the ToC?	Are contextual factors that might influence these pathways effectively integrated into the analysis?	Does the research report or evaluation provide insights or recommendations for refining the ToC based on empirical evidence?	Does the evaluation employ an appropriate design (experimental, quasi-experimental, or robust observational) to test its causal claims?	Are the sampling methods and data collection techniques adequately described and justified?	Are the analytical methods clearly explained and transparently applied, with limitations or potential biases openly discussed?	In parallel, each study will be assigned a score on the Maryland Scientific Methods Scale (1–5), where: 1: Weak design (e.g., no valid counterfactual), 3: Moderately robust quasi-experimental design, 5: Strong design (e.g., randomized controlled trial)
1. The Cash Lab. (2022).	1 This report focuses on impact assessment of the program on earnings, savings, MPI and closing the living income gap. No ToC is articulated. Causal pathways are implied (e.g. when mentioning the VSLA as a first step towards financial exclusion) but not explicitly identified or problematized.		1 The report does not explicitly link measured outcomes to a ToC or hypothesized causal pathways. Contextual factors are minimally integrated into the analysis. Savings and alternative income generating activities could be considered as measuring intermediate outcomes of a causal pathway, but this is not theorized as such. The report provides some background on the cocoa industry and poverty issues (p.4) but does not analyze how specific contextual factors might influence program outcomes (i.e. it doesn't explore how certain characteristics of the location or program participants moderate the program's impact).				2 The report provides insights and recommendations that could inform refinements of the program or its underlying ToC, but they are not explicitly linked back to any ToC. A finding like the *worsening* dietary diversity theory (hypothesis: preference for unhealthy food, p. 10) could be used for program refinement.	3 The research applies an appropriate design (quasi-experimental). The main limitation of this quasi-experimental design is that it doesn't rule out other relevant factors causing the observed 'treatment effect', for example initial land ownership (or other onobservables). Despite these caveats, the results still indicate program impact. As there is no pre-measurement (treatment-endline is compared to an (ex-post) control group, formed at endline), no initial balance of groups can be presented. They do reflect upon this themselves and discuss potential biases openly (e.g. p.9). Methods are transparently applied. The sampling methods and data collection techniques are partially described but lack comprehensive justification.			2 The research is using adequate control variables (as covariates in matching) and a cross-sectional comparison of treated groups with an untreated group. Matching techniques are used to account for cross-sectional differences between treated and controls groups.
2. B. Kell. (2020).	4 This thesis zooms in on how economic capabilities lead to different dimensions of well-being. Well-being is not an outcome of explicit interest for 100Weeks/HWG. However, together with other theses (doc 6+7), it informs the ToC, because well-being is hypothesized to impact economic capabilities. Taking together that a) this report has an elaborate ToC of its own with explicit causal pathways (e.g. greater instrumental freedoms lead to greater wellbeing, p. 26) and b) this, in the end, informs the ToC relevant to 100W/HWG, the score is 4.		3 The thesis explicitly links its findings (qualitative outcomes related to cognitive and relational well-being) to its own conceptual scheme (its own ToC), addressing the hypothesized pathways from economic capabilities to well-being (p. 74, Figure 14; p. 84, Figure 15; p. 87). For example, it links improved marital relationships and reduced loneliness (relational well-being) and increased happiness and aspirations back to the economic stability and empowerment provided by the program (e.g. p. 71, 78, 82). The ToC is adapted based on empirical evidence (e.g. p. 71, figure 13). Contextual factors influencing the outcome are identified and discussed (p. 38-45). A limitation is the lack of systemic testing with evidence of significance (e.g. results are substantiated by showing quotes) that stems from the qualitative method.				4 Its own conceptual framework — basically functioning as the ToC of 100Weeks — is refined based on this research (e.g. barriers are added in the framework). This is also carefully reflected upon. Also, relevant policy recommendations (p. 92-93) are given (e.g. ultra-poor are not reached, try other targeting aimed at including socially excluded ultra-poor).	2 The thesis uses an appropriate design within the qualitative domain and sampling is well-described (p. 30-34). The research provides an elaborate explanation of potential biases (with links to its epistemological stance). The researcher is talking to non-beneficiaries as well and summarizing that, e.g. in table 9, leading to a treated-control comparison. However, sample size is small (N-treated=11 and N-control=17) and their qualitative method is not equipping the researcher to make strong causal claims, especially not to people not included in the conducted interviews.			1 This thesis uses a cross-sectional comparison of treated groups with untreated groups.
3. Wolf et. al. (2024).	4 The paper outlines two different possible models that both function as a ToC and elegantly compares the two. Additionally, the paper zooms in on the question: 'do the cash transfers lead to more educational aspirations/engagement?' Which is an important step in between		4 The paper tests underlying assumptions of the 100Weeks program (cash transfer → less maternal stress, from the Family Stress Model; and cash transfer → more educational aspirations and engagement, from the Family Investment model). They searched for heterogeneity to check for mediators/contextual factors.				2 Recommendations for future research are given. However, actionable insights or recommendations for the 100Weeks program are not provided.	4 Yes, the evaluation employs a strong community-randomized controlled trial (RCT) design, which is highly appropriate for testing causal claims regarding the intervention's impact. Randomizing at the community level helps manage contamination within communities (spill-over effects). Standard errors are clustered as they should. The paper adequately describes the sampling method (p.3-5)			5 This research follows the 'gold standard' with respect to generating causal evidence in economic research. Baseline equivalence of groups before treatment

	the cash transfers and the 100Weeks program. Furthermore, they measure maternal stress, which is also an hypothesized pathway of program impact.			and data collection techniques (e.g. specific validated scales used for stress and economic indicators). The analytical methods are clearly explained and appropriate for the RCT design. imitations are openly discussed, including the midline nature of the results, reliance on self-reported data (e.g. potential social desirability bias, p.7).	is shown (p. 6-7) and control variables are used to correct for initial differences, but the main results are not impacted (p. 8).
4. M. Saleh (2024).	3 The thesis compares two competing models to describe household dynamics (cooperation model vs non-cooperation model), which gives us insights into the causal mechanisms of the 100Weeks model. However, the thesis does not go into detail about the 100Weeks program and its subcomponents.	2 The thesis does link some relevant outcomes (e.g. income sources, decision-making variables) to the ToC, but a major shortcoming is that the causal mechanisms of the ToC are not examined to a large extent, as the available data does not allow for that. Quote from report, p. 51: "This study could not provide a clear understanding of the underlying mechanisms by which cash transfer programs lead to increased income and household bargaining power."	1 The thesis provides general recommendations for policy and future research (p. 49-52). However, it offers no actionable recommendations for refining the program's design or its underlying ToC (e.g. to address the negative impact on autonomy over earnings)	3 The study design is generally appropriate as this thesis uses a subsample of data from the RCT-study (document 3). However, the study fails to clearly explain how the control group and treatment group differ, which is: they both engaged in VSLA, but the treatment group also got weekly cash transfers. The models presented have low explanatory power. Reflection on this is given (p. 35; other unknown factors likely have major influence on outcomes) but the validity of the constructs are not properly discussed. That is a major shortcoming as this could also be a reason for the low explanatory power of the models. Generally, the construct-validity is not always convincing (e.g. income hiding measured by women's contributions to the household).	4 As this thesis is using a sub-sample of the RCT-data from document 3 but no attention is paid to potential bias due to subsample criteria (and impact on power of the study), the score is lowered from 5 to 4.
5. J. M. Anne (n.d.).	3 This thesis zooms in on how life skills training and financial literacy influence different forms of capital (e.g. financial, human, relational, etc.), which leads to e.g. aspirations, beliefs and empowerment, which in turn will lead to increased income, improved skills and more. This is relevant to the causal mechanisms of the program. Pathways to the presented outcomes (p. 16-26) lack detail and clarity.	3 The thesis tests core pathways of the ToC that are underlying the program impact, and the contextual factors are integrated. Qualitative findings regarding changes in women's various capitals (assets) and personal agency are linked back to the pathways outlined in its SL-DP theoretical framework (its research ToC) (p. 69-90). It also considers how factors like initial health, household structure, and pre-existing vulnerabilities influence women's ability to benefit from the program (p. 97), thus integrating contextual factors in the analysis. A limitation is the lack of systemic testing with evidence of significance (e.g. results are substantiated by showing quotes) that stems from the qualitative method.	3 Recommendations for refining the ToC (the SL-DP framework application) and for 100Weeks are given (e.g. give participants the opportunity to give feedback on the trainings, p. 104-105), but many recommendations are not substantiated or clearly explained.	2 The method of this study is not designed to rigorously test causal claims about the average treatment effect compared to a counterfactual group. The observational study is using triangulation to substantiate findings. The convenience non-probability sampling used causes potential bias (p. 32). The proclaimed randomness of the simple probability sampling for past program participants is not adequately explained.	1 This thesis explores self-reported program effects on treated individuals only.
6. A. Langener (n.d.).	3 This study tests the feedback loop/bi-directional relationships between income, wellbeing and empowerment. This is embedded in a ToC, but this lacks detail (p. 10-11, Figure 1). The specific nature and role of the training component are not detailed.	2 Pathways from the intervention itself to each outcome variable are tested separately. However, the thesis does not explicitly state or directly test the underlying theoretical assumptions about why income, social support, wellbeing, and empowerment are expected to be related (e.g., assumptions drawn from psychological or economic theories). The finding of weak and unexpected relationships implicitly challenges some common theoretical assumptions about direct, strong links between these constructs, but the thesis doesn't frame this as a formal test of those assumptions. Contextual factors are not well integrated.	2 Recommendations for future research are given (p. 29-32). However, actionable insights or recommendations for the 100Weeks program are not provided.	2 The study design does not allow for credible causal inferences and SEM is not complemented with arguments (like Temporal Precedence) that support causality. The negative effect of empowerment on wellbeing is unexpected and unexplained, the 0-effect of income on wellbeing is not explained with power calculations or other hypotheses. Their measured variables might not cover the complex constructs well (quote, p. 29: "empowerment was only measured with two questions in the full sample"). These limitations are openly discussed but not to the full extent.	1 There is a 'control group', which is another group at baseline. A time trend is included as robustness check (p. 43) no other controls for initial differences or matching techniques are included.
7. W.H. Hijmissen-Bonhof. (2020).	2 The relationship between financial shocks and mental health is relevant for the core program pathway of cash	2 The thesis explicitly links its measured outcome (changes in mental health, measured by GHQ-12) to its primary variable of interest (various measures of financial	2 The income shocks effect on mental health (more money → more	3 The thesis properly describes the analytical method (fixed-effect regression models on panel data) and potential biases and limitations (e.g. p. 30-31; no control for food insecurity	1 This study looks at variation within the group of beneficiaries,

	transfers to mental health (p. 6-10). However, hypothesized causal pathways to different program components are not explicitly identified.	shocks), directly addressing the hypothesized pathway that financial instability affects mental wellbeing (p. 2; p. 20-26). While statistical links are established, the specific psychological or social mechanisms driving this link (e.g., stress pathways, coping mechanisms) are discussed theoretically (p. 6-10) but not empirically tested within the analysis. Within its own research question, it doesn't deeply explore intermediate mechanisms between the shock and the mental health outcome (e.g., changes in perceived stress, coping strategies employed).	mental health) is confirmed. No further patterns clearly emerge regarding income fluctuations and mental health. This thesis does not link this back to policy recommendations.	and confounding effect of general emotional support due to the program). This design is appropriate for its specific research question as it controls for time-invariant unobserved individual characteristics that could confound the relationship. Critically, this design cannot establish the causal effect of the 100Weeks program itself compared to non-beneficiaries. The sampling method (original sampling strategy of data-collectors) is not well-explained.	which is not comparing treated and untreated individuals with the intention of obtaining an estimate of a treatment effect.
8. A. Langener (2020).	2 The ToC (p. 6) lacks detail and is not clearly linked to the intervention components separately. Yet, expected outcomes are relevant to the program mechanisms and they are listed and categorized into first order and second order.	2 The thesis measures impact on key components that are steps in the program outcome (consumption, assets, savings, labor supply). Contextual factors are included in the heterogeneity analysis. The analysis does not provide a comprehensive assessment of all the relevant assumptions and causal pathways relevant for the ToC.	2 Recommendations for future research are given (p. 41). However, actionable insights or recommendations for the 100Weeks program are not provided.	2 The inclusion of many controls at once in the heterogeneity analysis (p. 47) complicates the interpretation of coefficients (what do they mean, what do they capture?). Potential biases are mentioned but not well integrated in the analysis (e.g. p. 41). There is some attempt to form a control group, who started later with the program (1 year later), but the exact comparison made and the rationale in this is not clearly explained. Results are presented transparently, including robustness checks across different model specifications (p. 19-34 & Appendix).	2 The pre-post comparison is corrected for macro level factors (inflation, a nested data structure, autoregressive effects, and calendar fixed effects). A kind of control group is formed, but there is no baseline equivalence or corrections based on initial differences.
9. L. Klunder (n.d.).	3 The thesis develops a relevant ToC on the topic of empowerment but lacks detail on all different program components. The research adopts the established Three-Dimensional Women's Empowerment Model (personal, relational, environmental) as its analytical framework to assess the program's impact. Hypothesized causal pathways link components of this combined approach to positive impacts across the three dimensions of empowerment. For example, environmental (societal) empowerment is linked to building a connection with peers in the saving groups (p.10-12).	2 The study examines outcomes across the three broad dimensions of empowerment (personal, relational, environmental), which themselves can be considered crucial intermediate outcomes on the path to broader goals like poverty reduction or enhanced well-being. Critically, while it measures these dimensional outcomes, the analysis does not delve deeply into the specific mechanisms of change within each dimension (e.g. how did training influence self-esteem exactly? What specific relational dynamics were altered by the cash?)	2 Recommendations for future research are given (p. 29). However, actionable insights or recommendations for the 100Weeks program are not provided.	3 This thesis follows a quasi-experimental design using Propensity Score Matching with endline data compared to a group at baseline that serves as control. This is an appropriate design but does not produce causal evidence. The thesis includes a good discussion of limitations, acknowledging the cross-sectional data limitations, the potential for omitted variable bias inherent in PSM (selection on unobservables), and potential measurement error. The inclusion of sensitivity analysis adds rigor to the PSM approach.	2 The research is using adequate control variables and a cross-sectional comparison of treated groups with an untreated group. Matching techniques are used to account for cross-sectional differences between treated and controls groups.

Table 20 MPI comparability and decisions for reduced MPI construction

Topic	100Weeks	RTV	MPI decisions
Standard of living			
Assets* (MPI)	Radio	Radio	Included 3 common assets and 2 items: wheelbarrow and machete. Used expenditure to proxy ownership of vehicles.
	TV	TV	
	Phone	Phone	
	Motorbike	Expenditure on vehicles	
	Fridge	-	
	Truck	Expenditure on vehicles	
Sanitation*	Toilet type in use	Toilet type in use	Harmonized and converted to 0/1
Drinking water*	Type of water source	Type of water source	Harmonized and converted to 0/1
	Travel time to source	Travel time to source	
Electricity*	Electricity access	Electricity	Harmonized and converted to 0/1
Housing*	Major floor material	Major floor material	Harmonized and converted to 0/1
	External wall material	External wall material	
	Roof material	Roof material	
Fuel for cooking*	Main fuel type used for cooking	Main fuel type used for cooking	Harmonized and converted to 0/1
Education			
Children school attendance*	School attendance of children	School attendance of children	Harmonized and converted to 0/1
Schooling years*	Completed years of education of each household members	Completed years of education of household head	Used information on household head only instead of all members.
Health			
Nutrition*	HFIAS-questionnaire that covers food insecurity	-	Modelled HDDS thresholds for HFIAS categories using ordered logistic regression. Predicted at risk in RTV dataset. (correlation -0.5)
Nutrition	Food diversity score (HDDS)	Item level consumption (can be converted to HDDS)	
Child mortality*	Under-5 mortality in last 5 years	-	Omitted entirely from reduced MPI

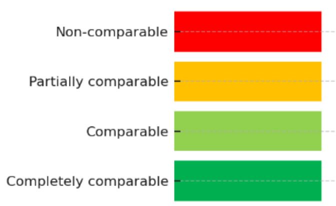


Figure 25 Daily per capita consumption (predicted) in 2017 PPP USD for the 100Weeks dataset by MPI score category for BL

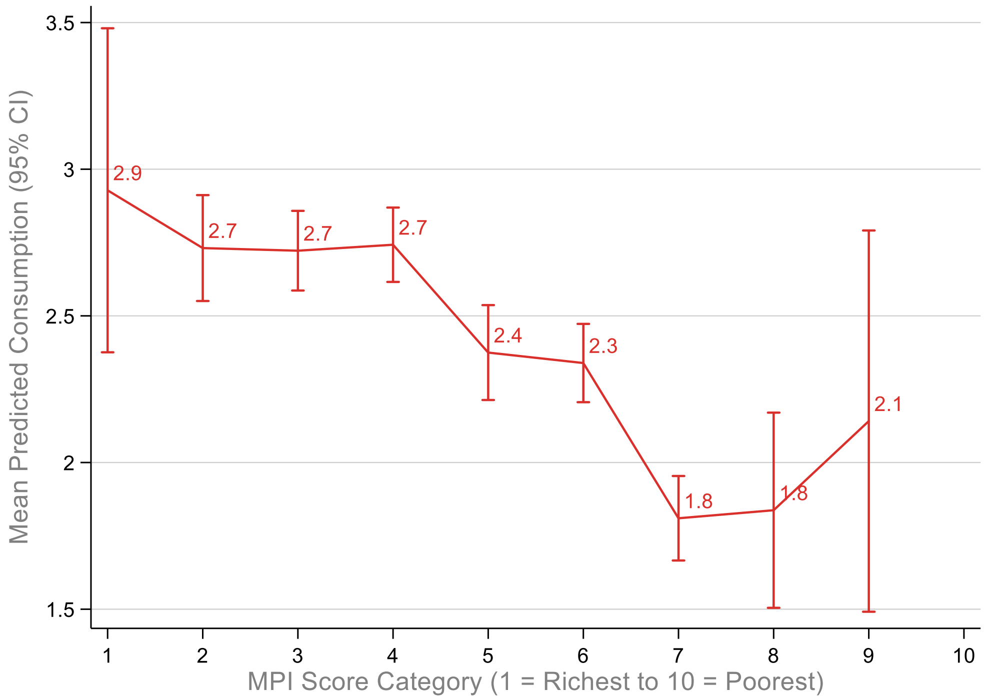
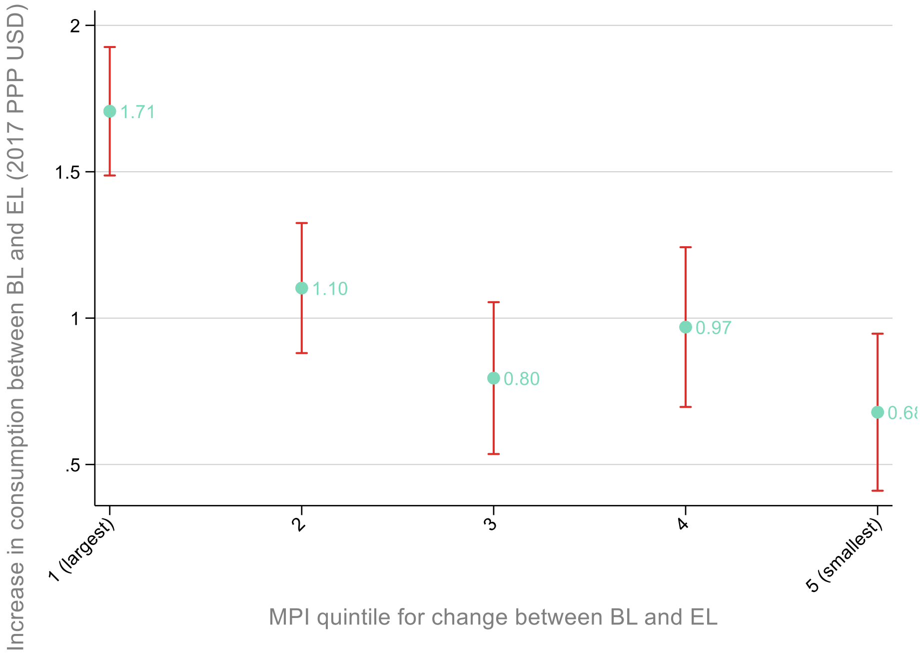


Figure 25 shows the baseline (Round 0) predicted household consumption across ten MPI score categories, ranging from 1 (Richest) to 10 (Poorest). Each point represents the mean predicted consumption, with vertical lines indicating the 95% confidence interval.

The graph shows a clear downward gradient in predicted consumption as MPI deprivation increases. The top three MPI categories (1–3) show similar consumption levels around 2.9–2.7 PPP 2017 USD, while households in the poorest MPI categories (7–9) fall to levels near or below 2.0. This suggests that MPI categories, constructed from non-monetary indicators, track well with predicted economic well-being. The confidence intervals widen in the poorest groups, reflecting smaller sample sizes (i.e. the last group 9 contains only 3 observations)

Figure 26 shows the relationship between improvements in multidimensional poverty (MPI) and increases in consumption between baseline and endline. Households in the first quintile- those experiencing the largest reductions in MPI, saw the highest consumption gains, with an average increase of 1.71 PPP USD per day. In contrast, those in the fifth quintile-households with the smallest changes in MPI, saw the lowest consumption gains, averaging 0.60 PPP USD. The pattern across quintiles shows a clear and approximately monotonic gradient: larger reductions in MPI are associated with greater improvements in consumption.

Figure 26 Increase in predicted per capita daily consumption per MPI change between BL and EL quintile (1 experienced largest change in MPI, 5 the smallest)



laterite

DATA | RESEARCH | ANALYTICS

From data to policy

www.laterite.com