Microfinance's Transformational Potential: Looking Beyond Average Treatment Effects*

Ronald A. Cueva[†] Adam Osman[‡] Jamin D. Speer[§]

August 2023

Abstract

Hundreds of studies have examined the impacts of microfinance, finding mostly modest or disappointing results. In this chapter, instead of asking whether microfinance works on average, we study the varied impacts of microfinance. Using data from several prominent recent studies, we show that the heterogeneity in returns to microcredit, microsavings, and microinsurance is large. This means that even programs that are not effective on average could be transformational for some people. We call for researchers and policymakers to focus more on identifying those who will benefit from microfinance, and understanding why they do. Together this will improve the targeting of these interventions, increase their positive impact, and help improve the design of future products.

*The authors thank Mohamed Meki & Simon Quinn for editing this volume, an anonymous referee for very helpful comments, as well as the other authors in the volume who provided feedback during the process. We also thank the authors of the studies used in our analysis for making their data publicly available and answering any questions we had. We also thank all of the individuals in these studies for their participation. All errors are our own.

[†]University of Illinois Urbana-Champaign, rcuevac1@illinois.edu

[‡]University of Illinois Urbana-Champaign, J-PAL MENA, and IZA, aosman@illinois.edu

[§]University of Memphis and IZA, jspeer@memphis.edu

1 Introduction

Does microfinance "work"? This question has driven researchers and policymakers for decades and has set sail to hundreds of research papers in economics and related disciplines. This plethora of studies has produced many answers to the question of whether microfinance works, not all of which agree. Some studies have found that microfinance can help increase income and savings and promote entrepreneurship, while other studies have found limited impacts on important outcomes and even negative effects in some cases.

The differences in findings across studies can be attributed to a variety of factors, such as variations in analytical techniques, disparate data sources from diverse locations and time periods, and comparisons of microfinance products with different features. But at their core, the vast majority of studies focus on comparing the average outcome from a group that got microfinance to the average outcome from a group that did not. For the cleanest comparisons, the two groups are randomly (or quasi-randomly) assigned so that their outcomes can be compared and interpreted as causal effects of offering the microfinance product. These effects are known as average treatment effects (ATEs).

While conclusions differ across studies and contexts, the general consensus is that existing studies on the effectiveness of microfinance have "failed to find transformational effects on key outcomes such as profits and income" (Cai et al., 2021). In Banerjee et al. (2015)'s review of the literature, they note a "consistent pattern of modestly positive, but not transformative, effects". Others are more negative, concluding that, based on the available evidence, "microcredit has not contributed significantly to poverty reduction" (Katreniak et al., 2022) or that "microfinance has been proven to have little or no positive impacts on people's lives" (Dahal and Fiala, 2020). In other words, the ATEs of microfinance interventions are typically small and often statistically insignificant. The popular press has sometimes not been so measured, with some labeling microfinance a "huge disappointment around the world" (Munir, 2014).

However, several recent papers have sought to broaden this narrative by looking at heterogeneity in returns to microfinance (Bryan et al., 2023; Crépon et al., 2022; Meager, 2019). That is, does access to microfinance have different effects for different people, and can we use this information to better design and target microfinance products? If some people benefit from microfinance, even if the average impact is negligible, then identifying who benefits and targeting them with future products can dramatically improve the average effectiveness of microfinance.

In this chapter, we make the case that a focus on average effects hides crucial insight about the varied impacts of microfinance. Building on these recent papers and employing methods used in Buhl-Wiggers et al. (2022) to measure the heterogeneity in the impacts of interventions, we use data from a number of prominent recent studies of microfinance products to show that the variance in treatment effects is very high, even under very conservative assumptions. We estimate bounds on the standard deviation of the estimated impacts and find that even the lower bound is large. That is, treatment effects can range from being very positive for some users to quite negative for others. This means that the focus on the average impacts of these products is only revealing a small part of the story. We show these insights to be true for the three main pillars of microfinance: credit, savings, and insurance.

There are two main takeaways from these exercises. First, even when a financial product is not effective on average, there is a good chance that the product is transformational for some people, with some people benefiting greatly and others being harmed. Similarly, programs that seem to work well on average do not work for everyone (Heckman et al., 1997). Second, instead of focusing on finding the best product designs, researchers and practitioners would benefit from spending more time figuring out how to better target the products we already have to those that benefit most from them. Current advances in machine learning provide one potential avenue for improving our product targeting capabilities (e.g., Chernozhukov et al. 2022), and we provide additional suggestions at the end of this chapter.

Our results imply that asking "Does microfinance work?" is not the best question for us to consider. A better question is, "For whom does microfinance work?" We show across a set of different microfinance experiments that there are some people who benefit a great deal from the product they gain access to. This can help reconcile the experience of practitioners who may have seen with their eyes how the product they provided was transformational in some people's lives while reading research that shows only modest average impacts from products like theirs.

2 Data

For our empirical analysis, we consider the data from six prominent recent studies of microfinance and ask what we can learn from them regarding the heterogeneity in the returns to these products and opportunities. Here, we will briefly describe the six studies, and then we will explain our empirical analysis.

Using the data from these studies, we analyze the impacts of three types of microfinance interventions: microcredit (loans), microsavings, and microinsurance. Microcredit is the most common type of microfinance, featuring small loans that must be paid back with interest. By microsavings, we mean expanding access to conventional bank accounts for individuals in developing countries, who traditionally lack such access. Microinsurance involves providing access to insurance against common risks, such as insufficient rainfall, in areas where such insurance is not commonly accessible to farmers or other entrepreneurs.

The six papers that we consider were chosen for a few reasons. We did not attempt a comprehensive or representative search of the microfinance literature. Instead, we looked for a manageable set of well-regarded papers using randomized controlled trials (RCTs) with randomization at the individual level and where the data were publicly available. Furthermore, we wanted studies that represented the three primary areas of traditional microfinance: credit, savings, and insurance. The papers we chose also cover a number of different countries and settings, ensuring that we are not getting results from only one part of the world or one type of country. One of the authors of this study also worked on the Crépon et al. (2022) and Bryan et al. (2023) papers, so those were chosen also for familiarity and ease of use.

Our first papers focus on access to microcredit, or loans to be paid back with interest. The oldest paper in our sample is Karlan and Zinman (2010). This paper uses a randomized experiment to study the impact of expanding access to expensive credit – that with a high interest rate – to borrowers in South Africa. Borrowers were randomly offered a short-term loan at an annual interest rate of 200%. The authors find that the expanded access to credit significantly improved borrowers' well-being on average, increasing their employment, income, food consumption, and a measure of subjective well-being. They also find that about two years later, borrowers were more likely to have a credit score and did not have lower credit scores than the control group.

On the other hand, Augsburg et al. (2015) study the impact of microcredit in Bosnia and Herzegovina using a standard randomized setup. The loans were meant to be for business activities, although this was not enforced or monitored, and the interest rate was 22%. They find that while the loans increased levels of business activity and rates of self-employment, they did not translate into increased household income. The authors suggest that the loans were too small to allow borrowers to effectively start or expand a business.

Crépon et al. (2022) use an RCT in Egypt to compare the impacts of microloans, cash, and in-kind grants relative to getting nothing; we will use data from this study to look at the impacts of traditional microcredit. They find that while both the loans and grants have similar positive effects on income, there is a large amount of heterogeneity within treatment groups. In particular, all three treatments had large benefits for a group at the top of the distribution and little impact for most recipients.

Instead of the normally small loans that are used in microfinance, Bryan et al. (2023) ask whether substantially larger loans can have larger positive effects on borrowers' businesses in Alexandria, Egypt. They randomly provided prior borrowers a loan that was four times as large as their previous loan. The authors find that the larger loans had small average impacts. However, the "top performers" – those with the largest treatment effects – saw substantial increases in profits, while those at the bottom of the distribution saw their profits actually fall. The results imply that there is a great deal of heterogeneity in who benefits (or even suffers) from getting loans, and that being able to effectively target loans can greatly improve the effectiveness of programs. As such, this is one of the papers that originally inspired the analysis for this chapter.

To look at the impacts of microsavings, we use data from Dupas et al. (2018), who study the impact of receiving access to basic bank accounts in Uganda and Malawi.[1] In both countries, the treatment groups that received access to basic bank accounts did not have higher savings or score higher on any other outcomes on average in comparison to the control groups.

Finally, we look at the impacts of providing microinsurance. Karlan et al. (2014)

[1] Dupas et al. (2018) also tested this impact in Chile, but they did not conduct follow-up surveys there due to a low take-up of the intervention in that country.

use a multi-year RCT to study the effect of providing rainfall insurance to farmers in northern Ghana. In this context, a year with inadequate rainfall can be devastating for small farmers and, on top of that, insurance against such events is difficult to access. This is important because while microcredit and microsavings address possible binding constraints for entrepreneurs, there can still be massive risk for those whose livelihoods depend on events outside of their control. The authors show that demand for this type of insurance product is high and that access to the insurance leads farmers to invest significantly more in their farms and make riskier production choices in agriculture.

Results from these studies include several different important outcomes of interest. We consider consumption of the participant as an outcome of interest for all of our studies. In the case of micro-credit, we also focus on business profits. Moreover, for micro-savings, we look at the impacts on savings behavior, while for microinsurance we consider investment and gross income.

3 Methodology

Our goal is to use the data from these papers to assess the degree of heterogeneity in treatment effects. While estimating average treatment effects is straightforward since the papers all use RCTs – one can mostly compare the average in the treatment group to the average in the control group – assessing heterogeneity is more difficult because it requires knowing something about how each individual is affected by the treatment. This presents a fundamental challenge: because no individual appears in both the control and treatment group, we cannot actually observe each person's individual treatment effect.

However, we can proceed by relying on the technique used by Buhl-Wiggers et al. (2022), which uses a method to put both lower and upper bounds on the variance of treatment effects and compare that variance to the average treatment effect (these are sometimes referred to as "FH Bounds" based on the work of Fréchet (1951) and Hoeffding (1941)). Their paper documents a high level of heterogeneity in the impacts of a literacy program in Uganda. We adapt it to do the same for the studies on microfinance that we are considering.

We proceed as follows, using the publicly available data from each of the papers we described above. We start with a conservative assumption called "rank preservation" that will give us a lower-bound on the true level of heterogeneity in treatment effects. First, we sort the treatment and control groups separately on the outcome variable (e.g., profits, consumption, etc.). Second, we compare the outcome for the "top" person in the treatment group to the top person in the control group; this is "rank preservation" because we are implicitly assuming that the top person in the treatment group would have been the top person in the control group had he or she not gotten the treatment. This difference in outcomes is the top person's treatment effect.[2]

We do the same for every other position in the data, from top to bottom. This gives us an estimated treatment effect for each person pair in the data. The variance of these treatment effects gives a lower-bound for the true variance in treatment effects in the experiment. This is because any change in the matching between people in treatment and control would change the treatment effects for those people in a way that would weakly increase the variance in the estimated effects (see Cambanis et al. (1976); Heckman et al. (1997); and Buhl-Wiggers et al. (2022) for more detailed discussions).

Under "rank inversion", instead of comparing the top person in control to the top person in treatment, we reverse the order, comparing the bottom person in control to the top person in treatment, and so on. The implicit assumption here is that the person with the lowest outcome in treatment would have been the one with the highest outcome if he or she were in the control group. The rest of the process is the same as under rank preservation. This gives an upper-bound for the variance in treatment effects.

We generate bootstrapped confidence intervals for both our estimates of the average treatment effect and the standard deviation of the estimated individual treatment effects. In each of these cases we use 1,000 bootstrapped replications and report the 95% confidence intervals.[3]

[2]If the treatment and control groups are of different sizes, we expand the groups to be of equal sizes so that each individual in the control group has a corresponding individual in the treatment group. We then sort the individuals based on their outcome, match them to the corresponding person in the control group, take the difference between them and then collapse the data by person in the treatment group (i.e., the weighted average). In cases of stratified randomizations, we implement this procedure within each stratum.

[3]Since the standard deviation is bounded below by zero, our bootstrapped confidence intervals may be biased. There are ways to utilize a bias-corrected and accelerated bootstrapping procedure to attempt to adjust for this (Davison and Hinkley, 1997), but since our estimates are usually far from

In our main tables, we standardize the outcome variables we are interested in by subtracting out the average and dividing by the standard deviation. This makes it easier to compare impacts and variance across studies. In the Appendix, we also report results using the raw data, in local currency, without standardizing (Tables A1, A3, and A5). Additionally, we report a set of results where we take the inverse hyperbolic sine of the outcome in Tables A7-A9.[4] This final alternative is useful for thinking about how the treatments change people's outcomes. If the treatment works by increasing outcomes by a percent change instead of a flat amount (i.e., credit increases everyone's profits by 10% as opposed to increasing everyone's profits by $100), then we could detect heterogeneity in outcomes using the standardized measure that would mask a homogeneous percent effect that only differs because of differences in people's baseline level of profits.

To help better understand the practical implications of these measures of the variance, we include two other sets of estimates. First, we report the proportion of the sample that is estimated to have a positive treatment effect, which gives us a sense of the fraction of people who are benefiting from the financial product. Next, we take the treatment effects we estimated under the different ranking assumptions and sort them from largest to smallest. We then report the range of the treatment effects, showing the 5th percentile of effects, as well as the 25th, 50th, 75th and 95th percentiles of effects. Note that neither of these sets of estimates is subject to the bounds estimated from our main analysis, but we include them as a more concrete way to think about the heterogeneity in treatment effects.

In the Appendix, we additionally report how the estimated treatment effect varies depending on where the individual ranks relative to others in the sample at the time of the follow-up survey (Tables A2, A4, and A6). For example, in Table 1, when we assume rank inversion, we see that the 95th percentile impact on consumption is 1.03 standard deviations, while the 5th percentile effect is -1.23. In Appendix Table A2, we see that those with the 95th percentile of consumption *at the time of the follow-up survey* experience an increase of 0.05 standard deviation increase in their consumption,

the boundary, we think the potential bias in our case is small.

[4] We use the inverse hyperbolic transformation instead of taking the logarithm to simplify dealing with the zeroes that are in the data. If we were to do logs and restrict to only positive values of the outcomes, we get the same qualitative results.

while those at the 75th percentile experience an increase of 0.87 standard deviations. So it is not the case that the people with the most consumption are also the ones with the largest increase in their consumption. This alternate way of ordering the treatment effects could be useful when considering how to target these products to those with the highest impacts. We discuss this in more detail in Section 5 below.

It is worth noting that even our lower bound estimates are likely underestimates. This is because for many of these experiments, take-up of the products is far from universal; for instance, only about 25% of people make deposits in the Dupas et al. (2018) paper. This means that selection into using the product is endogenous; that is, only people who think this product is good for them will use it. So in our analysis we have many people who have not used the product, giving them a zero treatment effect and decreasing variance relative to if they had a negative treatment effect.

4 Results

Microcredit

The preceding exercise allows us to produce the data shown in Tables 1-3. Table 1 shows the estimates from the studies dealing with microcredit (loans). The top panel shows the treatment effects on monthly profits, while the bottom panel shows the treatment effects on consumption. We standardize our outcome variables so that the estimates can be compared across studies. The first row shows our calculated average treatment effect (ATE) from each study, followed by bootstrapped confidence intervals for that estimate. The second row reports the standard deviation of the estimated treatment effects, and its confidence interval.[5] The final row shows the fraction of estimated individual treatment effects that are positive. We follow this with the proportion of the sample that have a positive estimated individual treatment effect. Finally, we show the range of estimated individual treatment effects in the sample.

In Panel A, consider the left two columns, which use the profits data from the microloans in Crépon et al. (2022). The average treatment effect is 0.11 standard deviations, but the standard deviation in treatment effects – a lower-bound of the true

[5] Bootstrapped confidence intervals were calculated using 1,000 replications.

standard deviation in impacts – is 1.17. Using rank inversion to get an upper-bound of the heterogeneity in treatment effects, we get a value of 1.67. In other words, the standard deviation in treatment effects in this study is 10 to 15 times as large as the ATE. About one-fifth of the estimated treatment effects are positive, and as we explore the range of treatment effects further, we find that the 95th percentile of treatment effects is 1.18 standard deviations; the treatment effects at the 25th, 50th, and 75th percentiles are zero; but at the 5th percentile it is -0.67 standard deviations. Together this tells us that even though there is a positive average treatment effect, there is a great deal of heterogeneity, with transformatively positive results for a small group of people and potentially transformatively negative results for another group, with most people not affected at all. Previous studies have primarily focused on the ATE, but these results show that the ATE only tells a small part of the story.

This type of result is not unique to the Crépon et al. (2022) study. The columns marked "Bosnia" show results obtained using the data from Augsburg et al. (2015). We find a similarly sized average treatment effect, but under rank preservation the lower-bound shows a standard deviation of about 9 times the ATE (0.79 vs. 0.09), while the upper-bound from rank inversion (1.50) is about 17 times the ATE. The last row shows that 26% of the people in the study had a positive estimated treatment effect, meaning that their income was higher in treatment than it would have been in the control group. So while the ATE is near zero, a quarter of participants benefited from the loans, and some benefited a great deal. We find similar patterns for all the papers we consider in this section. For instance, for the profits data from Bryan et al. (2023), which we label as "Egypt (Alexandria)", the lower-bound for the standard deviation in treatment effects is about 3 times as large as the ATE, and the upper-bound is 25 times the ATE.

Panel B repeats the analysis for measures of consumption and finds even greater evidence of extensive heterogeneity. In this panel, we are able to include data from Karlan and Zinman (2010) in the "South Africa" columns, where they provided a consumer loan (while the previous papers were providing business loans). Even though there is an average treatment effect of 0.14 standard deviations, we once again see evidence of substantial heterogeneity in treatment effects, with a standard deviation of treatment effects that is three times the size of the ATE, and evidence that impacts are not positive for the majority of the sample.

What are the takeaways from this exercise? First, across a variety of studies of microloans in different countries with different samples and different products, we consistently find that there is a large degree of heterogeneity in impacts. The standard deviation of effects is always much larger than the average effect, even under the most conservative assumptions. The heterogeneity is not just part of the story; one could argue it is the main story that the data are telling us.[6]

Second, in every case we consider, there seem to be some people who benefit greatly from the microloan they are offered. Even when the average effect is negligible – generally considered a disappointment by the researcher – there are likely people who see huge gains from access to the product. For policymakers, this is tremendously important. Identifying these people and targeting them can turn a disappointing microfinance product into a wildly successful one. Instead of seeing microfinance as a disappointment overall from the average treatment effects, this analysis helps us see it as a challenge with huge possibilities for transforming businesses and lives. The challenge is to find the right people.

Microsavings

Table 2 shows the results for microsavings, using data from two experiments in Dupas et al. (2018). Panel A reports results focused on standardized savings outcomes, while Panel B considers standardized consumption outcomes. Columns 1-4 focus on the experiment in Uganda, while columns 5-8 focus on the experiment run in Malawi. Average treatment effects in both Uganda and Malawi on bank savings are positive, but we can see from the percentile estimates that this is driven almost entirely by a small fraction of those effects that are positive and very large. Most of the sample does not end up with higher bank savings, but a few people end up with much higher bank savings. The standard deviation in treatment effects here is about 4-6 times as large as the ATE.

When we consider impacts on total savings, which include savings outside of the

[6] Inspired by Banerjee et al. (2019), we have performed our analysis separately for those who owned businesses before the intervention and those who did not. We find similarly large standard deviations of treatment effects within each group, suggesting that the heterogeneity in treatment effects is not easily explained by variation in business experience across the sample.

banking system, we find much more muted impacts, with an ATE of around 0.05, and statistically indistinguishable from zero. Nonetheless the standard deviation of the impact is large, nearly ten times the size of the ATE in Uganda, and even larger in Malawi.

In Panel B, the results for consumption show a similar pattern. The average treatment effects are small, but the standard deviation of the treatment effect is large. Under both types of ranking assumptions, the proportion of the sample that experiences a positive effect is nearly half, which suggests that even with the small ATE, many people are likely benefiting from the intervention.

In this sense, we again see that for access to savings, the heterogeneity is perhaps a bigger story than the average effect. Yet while the results from loans showed that a product with a small ATE could help some people, the results from savings show that a product with a strong ATE (e.g., on bank savings) could be driven by a small subset of the sample and not actually benefit the majority of the sample.

Microinsurance

Table 3 looks at the effects of being offered rainfall insurance in Ghana as reported in Karlan et al. (2014). We focus on standardized measures of gross income and consumption. We consider two treatment arms, one that provides rainfall insurance only, and another that provides the insurance combined with a cash grant. The insurance increased gross income by 0.56 standard deviations on average, while the Insurance+Grant arm increased it similarly, by 0.45 standard deviations. This is a very exciting increase in income, but as with the other methods we see a lot of heterogeneity, with a standard deviation that is more than twice the size of the ATE, and with around one third of the sample not benefiting from the intervention.

We find the same pattern in the consumption data, which a strong ATE of an increase of 0.36 standard deviations for the insurance arm, but a smaller 0.04 increase for the Insurance+Grant arm. However, while the top of the distribution of impacts is very positive, the bottom of the distribution is quite negative. The standard deviation of treatment effects range between 1.03 and 1.73, or about 4-5 times the ATE in the insurance case, and two orders of magnitude larger in the Insurance+Grant arm.

Overall, the results in all the three cases – microcredit, microsavings, and microinsurance – arrive to similar conclusions: the heterogeneity in impacts of the microfinance product swamp the average effects.

5 Implications for Policymakers and Researchers

Our results show that the discussion surrounding the impacts of microfinance needs to be reframed. The simple "Does it work?" framework is inadequate. Rather, the appropriate question for microfinance is, "For whom does it work, and for whom does it not work?" Policymakers who may have the idea – based on previous research – that microfinance has disappointing overall effects should be encouraged by this reframing, as it means that there are many people whose lives have been meaningfully improved by these tools, and many more who we can help in the future.

Here, research and policy must work hand in hand. It is up to researchers to collect data and develop methods for identifying those who benefit from microfinance, and it is up to policymakers and practitioners to share their insights about who seems to benefit most from these programs and to find ways to target microfinance to them. This requires a new focus from researchers studying microfinance as well as open and clear communication between those researchers, policymakers, and practitioners.

For researchers, the large heterogeneity we have shown in impacts of microfinance requires a new focus in future studies. While the average treatment effect should, of course, always be reported and discussed, more attention should be paid to the variance of those treatment effects, and more effort should be focused on identifying the characteristics of those who have high treatment effects – that is, those who benefit from the microfinance products.

How can we identify who benefits?

There are several ways to try to find who benefits most from these types of products. The most basic is subgroup analysis. Researchers can divide their control and treatment samples into subgroups and calculate the treatment effects for each subgroup; then they can report which subgroups have the most positive effects. Some studies have done this. Banerjee et al. (2019), for example, define the participants in their sample who owned businesses before the intervention as "gung-ho entrepreneurs" and others as “reluctant entrepreneurs”. They find that the gung-ho entrepreneurs have large positive treatment effects from the intervention, while the reluctant entrepreneurs showed negligible impact. Meager (2019) looks at seven microfinance studies and finds something similar: those with prior business experience tend to have higher treatment effects than other participants. While this result is statistically significant, it is not consistent across all of the studies she considers. Crépon et al. (2022) show large positive impacts of loans for women in their sample, and muted impacts for men. Other studies have considered heterogeneity along other dimensions including education, age, employment, income, and financial discipline (Afzal et al., 2022) .

Subgroup analysis is a good start, but it may miss important interactions between group characteristics. For example, if it is true that females with the lowest household incomes benefit most from microfinance, but men with the lowest household incomes do not benefit, this will not be picked up by an analysis based solely on gender or household income. In theory, one can do the subgroup analysis with as many groups as one likes – for instance, cutting by both gender and household income – but this requires the researcher to guess ahead of time which characteristics might predict high treatment effects. It can also open the researcher up to accusations of data mining.

A more recent approach tries to harness machine learning techniques to estimate heterogeneous treatment effects and identify those who might benefit most. One promising advance is introduced in Chernozhukov et al. (2022), who propose a method for detecting heterogeneity using machine learning methods. In it, they build a model of what would happen to each person if they got the treatment and another model of what would happen if they did not get the treatment. They then subtract these two predictions from each other to estimate a “Predicted Individual Treatment Effect”. Critically, they build these models using only half of the data (the training set), and then validate them on the other half of the data (the testing set). Since the final estimates are not based on data coming from the individual themselves, the estimates are considered “honest” and when combined with correct statistical procedures they are able to find “real” heterogeneity and are not merely data mining.

Machine learning also has the advantage over classical subgroup analysis because it is able to account for important interactions between observable characteristics. In the example we used before, subgroup analysis will have difficulty identifying the group with the highest returns if that group is the intersection of, say, gender, education, and household income. Some machine learning methods attempt to predict treatment effects in non-linear ways which allows for greater flexibility in identifying useful heterogeneity. It can find the combination of characteristics of those with the highest predicted returns even if those combinations are difficult to guess ahead of time. While these methods may seem intimidating to some researchers and policymakers at first, some statistical packages allow for relatively easy implementation.

Finding Heterogeneity May Depend on Collecting New Types of Data

Both subgroup analysis and machine learning methods rely on the type and quality of the data that are collected about the participants in the study. If the researcher only has basic data about the participants – e.g., ethnicity/race, gender, marital status, education, age, and household income – then they may struggle to detect differences in treatment effects that arise from differences in a characteristic that was not collected.

One way to see the limitations of the data that are currently being collected is to implement a "Total Variance Decomposition" exercise. In this exercise, we can estimate how much of the variance in treatment effects that we estimate can be explained by the baseline data that are collected on the participants. We implement this exercise in a two-step process. In the first step, we calculate the variance of the estimated "Individual Treatment Effect" (ITE), which we call the *total variance*. Next, we run a regression of the ITE on baseline characteristics for those in the treatment group and then use the coefficients to get a "predicted ITE". We then take the variance of the *predicted ITE* and divide it by the *total variance* in step one, and report that ratio as a percentage in Appendix Table A10.[7] We find that the data on observable characteristics that are normally collected (e.g., age, gender, education, etc.) can only explain a small amount of the variance that we identify. In most cases it can explain less than 10% of the variance, while in some cases, particularly in the Karlan et al. (2014) study, it can account for more of the variance.

This exercise leads to two conclusions. First, in certain cases existing data sources

[7] We use the variables that are included in the baseline balance tables from each study. This means that our estimates are lower bounds, since it is likely that the studies collect additional data that could be informative, and that interacting different variables could increase the explanatory power of the existing data.

could likely help us better understand the heterogeneity in treatment effects that we estimate. Second, in most cases the existing data are not enough. Thus, researchers must also get more creative and comprehensive with their data collection. The more kinds of data we have about participants, the better chance we have to identify the characteristics that predict treatment effects.

This can also help us better understand *why* these treatments have varied impacts on different people, which we could then use to better design products that address the insights gleaned from improving our models of how the benefits are accrued. In other words, finding heterogeneous effects can allow us to update our theory of change for how the intervention works and changes people's lives. As our theory improves, we can use that to design better interventions and products. Hence, more time and effort spent understanding heterogeneous impacts can lead to a positive feedback loop for researchers who are excited about designing new products and programs.

But what kind of new data could researchers collect? Bryan et al. (2023) show that psychometric data can be key in finding heterogeneous effects. They find that borrowers who exhibit traits related to over-confidence do worse with larger loans than those who do not. Despite collecting a trove of standard data at baseline (e.g., demographics, business characteristics, cognitive scores, financial literacy, etc.), they find that it is the inclusion of the psychometric data that makes the difference.

There is also evidence that using community knowledge may be able to help us predict treatment effects. Hussam et al. (2022) show that entrepreneurs have high-quality information about the returns to capital for *other* entrepreneurs and that this information is valuable in identifying high-return entrepreneurs above and beyond observable characteristics. They estimate that using this community knowledge in allocating grants to entrepreneurs could have tripled the returns to their cash grant program. While entrepreneurs can manipulate their reports to favor themselves, the authors show that there are ways to design the survey to elicit honest answers. These methods provide a promising way forward for researchers and policymakers in better targeting microfinance to high-return entrepreneurs.

Another intriguing idea from McKenzie (2018) is to ask participants themselves what they think their personal treatment effect is. It could be that people know more about their own possible benefits from microfinance than the researcher knows, and perhaps asking them directly can help us improve in our targeting of these products. McKenzie studies participants in a business plan competition experiment, in which the winners were awarded $50,000. Both control and treatment participants are asked what they think their business outcomes would be if their treatment status were reversed. He finds that people's expectations of their own treatment effects are generally inaccurate and unreliable. At least from this study — although it is a different context than microfinance – it does not seem that participants can predict their own treatment effects in a way that could help policymakers target these products. But this type of data may be useful in other contexts or might be predictive when combined with other types of data.

Conclusion

Both researchers and policymakers need to adjust their focus when measuring the gains from microfinance and implementing these interventions. The focus should be on identifying those who will benefit most from microfinance – because research shows that there are many people who may benefit greatly and others who may be harmed – rather than simply asking if it works on average. Researchers will need to be creative with their data collection and data analysis to identify these people and should work closely with policymakers and practitioners to push the field forward and maximize the benefits of the tools at our disposal.

References

Afzal, U., G. d'Adda, M. Fafchamps, S. Quinn, and F. Said (2022): "Demand for Commitment in Credit and Saving Contracts: A Field Experiment," NBER Working Paper.

Augsburg, B., R. De Haas, H. Harmgart, and C. Meghir (2015): "The impacts of microcredit: Evidence from Bosnia and Herzegovina," American Economic Journal: Applied Economics, 7, 183–203.

Banerjee, A., E. Breza, E. Duflo, and C. Kinnan (2019): "Can microfinance unlock a poverty trap for some entrepreneurs?" NBER Working Paper.

Banerjee, A., D. Karlan, and J. Zinman (2015): "Six randomized evaluations of microcredit: Introduction and further steps," American Economic Journal: Applied Economics, 7, 1–21.

Bryan, G. T., D. Karlan, and A. Osman (2023): "Big loans to small businesses: Predicting winners and losers in an entrepreneurial lending experiment," NBER Working Paper.

Buhl-Wiggers, J., J. T. Kerwin, J. Muñoz-Morales, J. Smith, and R. Thornton (2022): "Some children left behind: Variation in the effects of an educational intervention," Journal of Econometrics.

Cai, J., M. Meki, S. Quinn, E. Field, C. Kinnan, J. Morduch, and F. Said (2021): "Microfinance," VoxDevLit, 3, 26.

Cambanis, S., G. Simons, and W. Stout (1976): "Inequalities for E k (x, y) when the marginals are fixed," Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 36, 285–294.

Chernozhukov, V., M. Demirer, E. Duflo, and I. Fernandez-Val (2022): "Generic machine learning inference on heterogeneous treatment effects in randomized experiments," National Bureau of Economic Research Working Paper 24678.

Crépon, B., M. El Komi, and A. Osman (2022): "Is it who you are or what you get? Comparing the impacts of loans and grants for microenterprise development," Working Paper.

Dahal, M. and N. Fiala (2020): "What do we know about the impact of microfinance? The problems of statistical power and precision," World Development, 128, 104773.

Davison, A. C. and D. V. Hinkley (1997): Bootstrap methods and their application, 1, Cambridge university press.

Dupas, P., D. Karlan, J. Robinson, and D. Ubfal (2018): "Banking the unbanked? Evidence from three countries," American Economic Journal: Applied Economics, 10, 257–297.

Fréchet, M. (1951): "On correlation tables with given margins," 14, 53–77.

Heckman, J. J., J. Smith, and N. Clements (1997): "Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts," The Review of Economic Studies, 64, 487–535.

Hoeffding, W. (1941): "Masstabinvariante korrelationsmasse für diskontinuierliche verteilungen," Archiv für mathematische Wirtschafts-und Sozialforschung, 7, 49–70.

Hussam, R., N. Rigol, and B. N. Roth (2022): "Targeting high ability entrepreneurs using community information: Mechanism design in the field," American Economic Review, 112, 861–98.

Karlan, D., R. Osei, I. Osei-Akoto, and C. Udry (2014): "Agricultural decisions after relaxing credit and risk constraints," The Quarterly Journal of Economics, 129, 597–652.

Karlan, D. and J. Zinman (2010): "Expanding credit access: Using randomized supply decisions to estimate the impacts," The Review of Financial Studies, 23, 433–464.

Katreniak, D. C., A. Khazanov, O. Moav, Z. Neeman, and H. Zoabi (2022): "The Microfinance Disappointment: An Explanation based on Risk Aversion," Working Paper.

McKenzie, D. (2018): "Can business owners form accurate counterfactuals? Eliciting treatment and control beliefs about their outcomes in the alternative treatment status," Journal of Business & Economic Statistics, 36, 714–722.

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

Munir, K. (2014): "Microfinance has been a huge disappointment around the world," Business Insider.

Tables

Table 1: Heterogeneity in Microcredit

	Egypt (Qena)		Bosnia		Egypt (Alexandria)		South Africa	
	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion
Panel A: Standardized Profits	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Average Treatment Effect	0.11		0.09		0.07			
	(0.06,0.17)		(-0.07,0.24)		(-0.03,0.17)			
Impact Standard Deviation	1.17	1.67	0.79	1.50	0.19	1.77		
	(1.06,1.31)	(1.55,1.82)	(0.48,1.09)	(1.16,1.84)	(0.02,0.37)	(1.58,1.96)		
Proportion Positive	0.21	0.28	0.26	0.35	0.49	0.50		
Percentiles of Treatment Effect								
5th	-0.67	-1.91	-0.51	-1.74	-0.02	-2.49		
25th	0.00	0.00	0.00	-0.29	0.00	-0.63		
50th	0.00	0.00	0.00	0.00	0.00	0.00		
75th	0.00	0.18	0.05	0.42	0.09	0.68		
95th	1.18	2.31	1.20	2.08	0.35	2.72		
Panel B: Standardized Consumption								
Average Treatment Effect	0.02		-0.11		0.07		0.14	
	(-0.03,0.07)		(-0.20,-0.03)		(-0.03,0.17)		(-0.01,0.29)	
Impact Standard Deviation	0.81	1.85	0.67	1.25	0.24	1.90	0.45	1.80
	(0.76,0.89)	(1.79,1.92)	(0.42,0.92)	(0.83,1.68)	(0.07,0.40)	(1.73,2.07)	(0.30,0.59)	(1.68,1.91)
Proportion Positive	0.52	0.51	0.31	0.48	0.62	0.51	0.13	0.37
Percentiles of Treatment Effect								
5th	-1.23	-3.04	-0.56	-1.77	-0.04	-2.83	-0.11	-3.58
25th	-0.26	-1.10	-0.12	-0.52	-0.01	-0.76	0.00	-0.11
50th	0.02	0.02	-0.04	-0.03	0.01	0.02	0.00	0.00
75th	0.29	1.05	0.02	0.38	0.10	0.76	0.00	1.56
95th	1.03	2.82	0.14	1.32	0.26	2.98	1.45	3.51

Notes: To generate these estimates treatment and control groups are first ranked from highest to lowest and then matched based on their rank. Under rank preservation (odd columns) those with the highest values in treatment are matched to those with the highest values in control. Under rank inversion (even columns) those with the highest values in treatment are matched to those with the lowest values in control. We then subtract the control value from treatment to get an estimate of the "individual treatment effect". We then report the average of this value across the sample as the average treatment effect, the standard deviation of these values, and the fraction of them that are positive. Finally, we sort these values from largest to smallest and report the percentiles of the sorted treatment effects. Panel A reports results from a measure of standardized profits, while Panel B reports the results from a measure of standardized consumption. Egypt (Qena) data are from Crepon et al (2022), Bosnia data are from Augsburg et al (2015), Egypt (Alexandria) data are from Bryan et al (2023), and South Africa data are from Karlan and Zinman (2010). They do not collect business profits data in the South Africa study. Bootstrapped confidence intervals with 1000 replications in parantheses.

Table 2: Heterogeneity in Microsavings

	Uganda				Malawi			
	Bank Savings		***Total Savings***		***Savings Bank***		***Total Savings***	
	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Average Treatment Effect	0.28		0.05		0.26		0.05	
	(0.19,0.37)		(0.00,0.10)		(0.17,0.35)		(-0.01,0.10)	
Impact Standard Deviation	1.09	1.69	0.48	1.61	1.28	1.75	0.71	1.54
	(0.89,1.28)	(1.51,1.87)	(0.39,0.56)	(1.51,1.70)	(1.05,1.52)	(1.51,1.99)	(0.61,0.81)	(1.42,1.66)
Proportion Positive	0.34	0.36	0.63	0.53	0.24	0.26	0.43	0.52
Percentiles of Treatment Effect								
5th	-0.01	-0.36	-0.62	-2.68	-0.01	0.00	-0.56	-2.37
25th	-0.01	0.00	0.00	-0.42	-0.01	0.00	-0.02	-0.29
50th	-0.01	0.00	0.01	0.03	-0.01	0.00	0.00	0.01
75th	0.13	0.18	0.09	0.46	-0.01	0.04	0.06	0.36
95th	1.78	2.35	0.62	2.59	1.49	2.04	0.65	2.12
	Food Consumption				***Food Consumption***			
Average Treatment Effect	0.04				0.04			
	(0.00,0.09)				(-0.03,0.07)			
Impact Standard Deviation	0.39	1.75			0.48	1.76		
	(0.34,0.45)	(1.67,1.83)			(0.42,0.54)	(1.69,1.82)		
Proportion Positive	0.49	0.51			0.53	0.51		
Percentiles of Treatment Effect								
5th	-0.33	-2.80			-0.56	-2.88		
25th	-0.03	-0.81			-0.06	-0.74		
50th	0.00	0.01			0.01	0.02		
75th	0.11	0.84			0.14	0.79		
95th	0.41	2.70			0.69	2.68		

Notes: To generate these estimates treatment and control groups are first ranked from highest to lowest and then matched based on their rank. Under rank preservation (odd columns) those with the highest values in treatment are matched to those with the highest values in control. Under rank inversion (even columns) those with the highest values in treatment are matched to those with the lowest values in control. We then subtract the control value from treatment to get an estimate of the "individual treatment effect". We then report the average of this value across the sample as the average treatment effect, the standard deviation of these values, and the fraction of them that are positive. Finally, we sort these values from largest to smallest and report the percentiles of the sorted treatment effects. Panel A reports results from standarized bank savings and total savings, while Panel B reports the results from standarized food consumption. The Uganda and Malawi data are from Dupas et al. (2018). Bootstrapped confidence intervals with 1000 replications in parantheses.

Table 3: Heterogeneity in Microinsurance

	Gross Income				Consumption			
	Insurance		***Insurance + Grant***		***Insurance***		***Insurance + Grant***	
	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ATE	0.56		0.45		0.36		0.04	
	(0.42,0.7)		(0.32,0.58)		(0.20,0.52)		(-0.07,0.15)	
Impact Standard Deviation	1.20	1.59	1.12	1.42	1.34	1.73	1.03	1.40
	(1.13,1.43)	(1.48,1.81)	(1.09,1.35)	(1.37,1.64)	(1.19,1.64)	(1.55,2.03)	(0.90,1.26)	(1.25,1.61)
Proportion Positive	0.67	0.65	0.74	0.66	0.48	0.47	0.49	0.48
Percentiles of Treatment Effect								
5th	-0.98	-1.78	-0.64	0.26	-0.90	-2.27	-1.28	-2.40
25th	-0.20	-0.44	-1.24	1.27	-0.18	-0.46	-0.37	-0.56
50th	0.50	0.35	-0.18	2.48	0.07	0.06	-0.02	-0.02
75th	1.11	1.22	0.63	1.21	0.65	1.05	0.34	0.67
95th	2.83	4.09	0.44	-0.74	3.02	3.44	1.62	2.21

Notes: To generate these estimates treatment and control groups are first ranked from highest to lowest and then matched based on their rank. Under rank preservation (odd columns) those with the highest values in treatment are matched to those with the highest values in control. Under rank inversion (even columns) those with the highest values in treatment are matched to those with the lowest values in control. We then subtract the control value from treatment to get an estimate of the "individual treatment effect". We then report the average of this value across the sample as the average treatment effect, the standard deviation of these values, and the fraction of them that are positive. Finally, we sort these values from largest to smallest and report the percentiles of the sorted treatment effects. The left panel reports results from a measure of standardized gross income, while the right panel reports the results from a measure of standardized consumption. The Gross Income and Consumption data are from the first year of intervention in Karlan et al. (2014). Bootstrapped confidence intervals with 1000 replications in parantheses.

Appendix Tables

Table A1: Heterogeneity in Microcredit (levels)

	Egypt (Qena)		Bosnia		Egypt (Alexandria)		South Africa	
	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion
Panel A: Monthly Profits	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ATE	95		747		1538			
	(55,135)		(-316,1811)		(-540,3617)			
Impact Standard Deviation	988	1415	6712	12799	4279	38783		
	(895,1104)	(1327,1518)	(4133,9291)	(10394,15205)	(917,7641)	(35518,42047)		
Proportion Positive	0.21	0.28	0.26	0.35	0.49	0.50		
Percentiles of Treatment Effect								
5th	-563	-1611	-4136	-14836	-414	-54678		
25th	0	0	-1	-2458	0	-13893		
50th	0	0	-1	-1	0	-60		
75th	0	150	396	3591	1995	14828		
95th	1000	1956	10238	17742	7690	59693		
Panel B: Monthly Consumption								
ATE	53		-699		290		0.10	
	(-73,180)		(-1328,-70)		(-111,692)		(-0.01,0.22)	
Impact Standard Deviation	2043	4673	4249	7928	976	7822	0.33	1.32
	(1918,2255)	(4538,4843)	(870,7628)	(5466,10391)	(376,1575)	(7261,8383)	(0.21,0.44)	(1.23,1.41)
Proportion Positive	0.52	0.51	0.31	0.48	0.62	0.51	0.13	0.37
Percentiles of Treatment Effect								
5th	-3117	-7706	-3564	-11167	-149	-11655	-0.08	-2.63
25th	-658	-2796	-743	-3269	-56	-3112	0.00	-0.08
50th	45	51	-273	-178	38	65	0.00	0.00
75th	728	2666	108	2407	429	3132	0.00	1.15
95th	2604	7141	891	8341	1080	12252	1.07	2.58

Notes: To generate these estimates treatment and control groups are first ranked from highest to lowest and then matched based on their rank. Under rank preservation (odd columns) those with the highest values in treatment are matched to those with the highest values in control. Under rank inversion (even columns) those with the highest values in treatment are matched to those with the lowest values in control. We then subtract the control value from treatment to get an estimate of the "individual treatment effect". We then report the average of this value across the sample as the average treatment effect, the standard deviation of these values, and the fraction of them that are positive. Finally, we sort these values from largest to smallest and report the percentiles of the sorted treatment effects. Panel A reports results from a measure of profits, while Panel B reports the results from a measure of consumption. Egypt (Qena) data are from Crepon et al (2022), Bosnia data are from Augsburg et al (2015), Egypt (Alexandria) data are from Bryan et al (2023), and South Africa data are from Karlan and Zinman (2010). They do not collect business profits data in the South Africa study, nor do they make a direct measure of consumption available, and so we use the reported index. Bootstrapped confidence intervals with 1000 replications in parantheses.

Table A2: Sorting Microcredit Impacts by Outcome at Follow-Up Survey

	Egypt (Qena)		Bosnia		Egypt (Alexandria)		South Africa	
	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion
Panel A: Standardized Profits	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Percentiles of Treatment Outcome								
5th	0.00	3.26	0.00	3.62	-0.01	2.96		
25th	0.00	0.10	0.00	0.42	-0.02	0.69		
50th	0.00	-0.01	0.03	-1.49	0.02	0.02		
75th	-0.07	-0.31	-0.02	-0.82	0.09	-0.60		
95th	0.55	-0.80	0.93	-0.60	0.37	-2.38		
Panel B: Standardized Consumption								
Percentiles of Treatment Outcome								
5th	0.06	2.44	0.05	2.52	0.04	3.28	-0.13	3.43
25th	0.18	1.53	-0.06	0.41	-0.01	0.78	0.06	0.65
50th	0.10	-1.65	0.06	0.04	0.03	0.03	-0.04	-0.03
75th	0.87	-0.90	-0.13	-1.47	0.06	-0.70	0.00	-1.71
95th	0.05	-1.40	-0.12	-1.12	0.41	-2.65	0.00	-1.56

Notes: To generate these estimates treatment and control groups are first ranked from highest to lowest and then matched based on their rank. Under rank preservation (odd columns) those with the highest values in treatment are matched to those with the highest values in control. Under rank inversion (even columns) those with the highest values in treatment are matched to those with the lowest values in control. We then subtract the control value from treatment to get an estimate of the "individual treatment effect". We then report the average individual treatment effect for individuals in treatment whose outcome variable is at the Xth perctile of the sample distribution. Panel A reports results from a measure of standardized profits, while Panel B reports the results from a measure of standardized consumption. Egypt (Qena) data are from Crepon et al (2022), Bosnia data are from Augsburg et al (2015), Egypt (Alexandria) data are from Bryan et al (2023), and South Africa data are from Karlan and Zinman (2010). They do not collect business profits data in the South Africa study. Bootstrapped confidence intervals with 1000 replication in parantheses.

Table A3: Heterogeneity in Microsavings (levels)

	Uganda				Malawi			
	Bank Savings		***Total Savings***		***Bank Savings***		***Savings Total***	
	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ATE	8.8		3.9		3.9		1.6	
	(6.93,10.67)		(55,135)		(2.93,4.87)		(55,135)	
Impact Standard Deviation	34.3	53.3	38.2	128.9	19.3	26.4	23.0	50.2
	(30.2,38.5)	(49.1,57.5)	(31.5,45.0)	(122.4,135.4)	(16.9,21.8)	(23.9,28.9)	(19.4,26.6)	(46.9,53.5)
Proportion Positive	0.34	0.36	0.63	0.53	0.24	0.26	0.43	0.52
Percentiles of Treatment Effect								
5th	0	-11.3	-49.4	-214.9	0	0	-18.4	-77.3
25th	0	0	0	-33.6	0	0	-1	-9.6
50th	0	0	1	2	0	0	0	0
75th	4.2	5.7	6.9	37.1	0	1	2	11.8
95th	56.4	74.4	49.7	208.0	22.4	30.8	21.1	68.9
	Food Consumption				***Food Consumption***			
Average Treatment Effect	0.3				0.3			
	(-0.04,0.63)				(-0.18,0.46)			
Impact Standard Deviation	2.8	11.8			3.3	11.3		
	(2.3,3.1)	(11.5,12.3)			(2.8,3.6)	(11.1,11.9)		
Proportion Positive	0.5	0.5			0.5	0.5		
Percentiles of Treatment Effect								
5th	-2.3	-19.8			-3.8	-19.7		
25th	0	-5.7			0	-5.1		
50th	0	0			0	0		
75th	1	5.9			1	5.4		
95th	2.9	19.1			4.7	18.4		

Notes: To generate these estimates treatment and control groups are first ranked from highest to lowest and then matched based on their rank. Under rank preservation (odd columns) those with the highest values in treatment are matched to those with the highest values in control. Under rank inversion (even columns) those with the highest values in treatment are matched to those with the lowest values in control. We then subtract the control value from treatment to get an estimate of the "individual treatment effect". We then report the average of this value across the sample as the average treatment effect, the standard deviation of these values, and the fraction of them that are positive. Finally, we sort these values from largest to smallest and report the percentiles of the sorted treatment effects. The upper panel reports results from bank savings and total savings, while the lower panel reports the results from a measure of food consumption. The Uganda and Malawi data are from Dupas et al. (2018). Bootstrapped confidence intervals with 1000 replications in parantheses.

Table A4: Sorting Microsavings Impacts by Outcome at Follow-Up Survey

	Uganda				Malawi			
	Bank Savings		***Total Savings***		***Savings Bank***		***Savings Total***	
	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Percentiles of Treatment Outcome								
5th	0.01	1.64	0.02	1.71	0.00	3.88	0.00	2.90
25th	0.05	0.28	0.13	0.26	0.00	0.14	0.00	0.48
50th	0.18	-0.15	0.00	-0.06	0.00	-0.18	-0.01	-0.49
75th	0.51	-0.64	0.00	-0.70	0.50	-0.04	-0.01	-0.48
95th	1.02	0.00	0.33	-0.70	1.17	-0.64	0.21	-1.33
	Food Consumption				***Food Consumption***			
Percentiles of Treatment Outcome								
5th	-0.03	1.76			0.02	3.33		
25th	-0.03	0.59			-0.04	0.93		
50th	0.05	-0.09			-0.01	-0.28		
75th	0.00	-0.93			0.07	-0.62		
95th	0.16	-1.28			0.12	-1.53		

Notes: To generate these estimates treatment and control groups are first ranked from highest to lowest and then matched based on their rank. Under rank preservation (odd columns) those with the highest values in treatment are matched to those with the highest values in control. Under rank inversion (even columns) those with the highest values in treatment are matched to those with the lowest values in control. We then subtract the control value from treatment to get an estimate of the "individual treatment effect". We then report the average individual treatment effect for individuals in treatment whose outcome variable is at the Xth perctile of the sample distribution. The upper panel reports results from bank savings and total savings, while the lower panel reports the results from a measure of food consumption. The Uganda and Malawi data are from Dupas et al. (2018).

Table A5: Heterogeneity in Microinsurance (levels)

	Gross Income				Consumption			
	Insurance		***Insurance + Grant***		***Insurance***		***Insurance + Grant***	
	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ATE	359		313		190		23	
	(282,437)		(237,389)		(116,264)		(-40,86)	
Impact Standard Deviation	775	1022	781	992	713	923	594	807
	(759,885)	(995,1118)	(777,925)	(976,1116)	(674,830)	(875,1028)	(542,701)	(739,912)
Proportion Positive	0.48	0.47	0.74	0.66	0.56	0.54	0.49	0.48
Percentiles of Treatment Effect								
5th	-630	-1145	-863	-1488	-478	-1213	-737	-1387
25th	-129	-285	-4	-182	-94	-247	-213	-324
50th	321	223	307	180	39	31	-10	-13
75th	715	786	548	878	349	561	198	387
95th	1819	2633	1805	2378	1612	1838	938	1278

Notes: To generate these estimates treatment and control groups are first ranked from highest to lowest and then matched based on their rank. Under rank preservation (odd columns) those with the highest values in treatment are matched to those with the highest values in control. Under rank inversion (even columns) those with the highest values in treatment are matched to those with the lowest values in control. We then subtract the control value from treatment to get an estimate of the "individual treatment effect". We then report the average of this value across the sample as the average treatment effect, the standard deviation of these values, and the fraction of them that are positive. Finally, we sort these values from largest to smallest and report the percentiles of the sorted treatment effects. The left panel reports results from a measure of gross income, while the right panel reports the results from a measure of consumption. The Gross Income and Consumption data are from the first year of intervention in Karlan et al. (2014). Bootstrapped confidence intervals with 1000 replications in parantheses.

Table A6: Sorting Microinsurance Impacts by Outcome at Follow-Up Survey

	Gross Income				Consumption			
	Insurance		***Insurance + Grant***		***Insurance***		***Insurance + Grant***	
	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Percentiles of Treatment Outcome								
5th	-0.98	0.08	-0.64	0.26	-0.01	-0.15	-0.07	-0.12
25th	0.64	0.29	-1.24	1.27	-0.60	1.16	-0.04	0.12
50th	0.70	-1.10	-0.18	2.48	-2.81	-0.01	-3.68	0.68
75th	0.09	-0.57	0.63	1.21	-0.14	0.08	-3.91	-0.97
95th	1.91	-2.52	0.44	-0.74	4.89	0.09	-0.77	-2.78

Notes: To generate these estimates treatment and control groups are first ranked from highest to lowest and then matched based on their rank. Under rank preservation (odd columns) those with the highest values in treatment are matched to those with the highest values in control. Under rank inversion (even columns) those with the highest values in treatment are matched to those with the lowest values in control. We then subtract the control value from treatment to get an estimate of the "individual treatment effect". We then report the average individual treatment effect for individuals in treatment whose outcome variable is at the Xth perctile of the sample distribution. The left panel reports results from a measure of gross income, while the right panel reports the results from a measure of consumption. The Gross Income and Consumption data are from the first year of intervention in Karlan et al. (2014).

Table A7: Heterogeneity in Microcredit (IHS transformation)

	Egypt (Qena)		Bosnia		Egypt (Alexandria)		South Africa	
	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion
Panel A: Standardized Profits	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Average Treatment Effect	0.70		0.60		-0.04			
	(0.59,0.83)		(0.04,1.16)		(-0.28,0.21)			
Impact Standard Deviation	2.44	4.80	2.25	7.57	0.93	4.81		
	(2.38,2.63)	(4.73,4.93)	(1.68,2.82)	(7.22,7.93)	(0.48,1.37)	(4.47,5.15)		
Proportion Positive	0.22	0.28	0.27	0.35	0.49	0.50		
Percentiles of Treatment Effect								
5th	-1.23	-8.04	-0.60	-10.26	-0.13	-11.60		
25th	0.00	0.00	0.00	-8.49	0.00	-1.48		
50th	0.00	0.00	0.00	0.00	0.00	-0.01		
75th	0.00	5.70	0.13	8.88	0.10	1.42		
95th	6.67	8.27	7.62	10.48	0.20	11.45		
Panel B: Standardized Consumption								
Average Treatment Effect	0.04		-0.18		0.07		0.09	
	(0.00,0.07)		(-0.35,-0.01)		(-0.06,0.20)		(-0.01,0.18)	
Impact Standard Deviation	0.57	1.37	1.01	2.35	0.44	2.44	0.27	1.11
	(0.48,0.69)	(1.31,1.45)	(0.69,1.33)	(2.04,2.65)	(-0.07,0.94)	(1.99,2.89)	(0.18,0.37)	(1.04,1.18)
Proportion Positive	0.52	0.51	0.27	0.35	0.62	0.51	0.13	0.37
Percentiles of Treatment Effect								
5th	-0.70	-2.09	-0.60	-10.26	-0.07	-2.41	-0.07	-2.16
25th	-0.26	-0.91	0.00	-8.49	-0.02	-0.80	0.00	-0.08
50th	0.02	0.02	0.00	0.00	0.01	0.02	0.00	0.00
75th	0.25	0.87	0.13	8.88	0.08	0.79	0.00	1.01
95th	0.76	2.07	7.62	10.48	0.19	2.49	0.92	2.13

Notes: To generate these estimates treatment and control groups are first ranked from highest to lowest and then matched based on their rank. Under rank preservation (odd columns) those with the highest values in treatment are matched to those with the highest values in control. Under rank inversion (even columns) those with the highest values in treatment are matched to those with the lowest values in control. We then subtract the control value from treatment to get an estimate of the "individual treatment effect". We then report the average of this value across the sample as the average treatment effect, the standard deviation of these values, and the fraction of them that are positive. Finally, we sort these values from largest to smallest and report the percentiles of the sorted treatment effects. Panel A reports results from a measure of inverse hyperbolic sine (I.H.S.) transformation for profits, while Panel B reports the results from a measure of I.H.S. transformation for consumption. Egypt (Qena) data are from Crepon et al (2022), Bosnia data are from Augsburg et al (2015), Egypt (Alexandria) data are from Bryan et al (2023), and South Africa data are from Karlan and Zinman (2010). They do not collect business profits data in the South Africa study. Bootstrapped confidence intervals with 1000 replications in parantheses.

Table A8: Heterogeneity in Microsavings (IHS transformation)

	Uganda				Malawi			
	Bank Savings		***Total Savings***		***Savings Bank***		***Total Savings***	
	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Average Treatment Effect	0.97		0.24		0.59		0.16	
	(0.89,1.05)		(0.14,0.33)		(0.52,0.64)		(0.07,0.24)	
Impact Standard Deviation	1.57	2.24	0.64	3.60	1.27	1.74	0.75	3.05
	(1.50,1.63)	(2.16,2.32)	(0.57,0.72)	(3.55,3.65)	(1.19,1.32)	(1.65,1.80)	(0.67,0.82)	(2.99,3.09)
Proportion Positive	0.34	0.36	0.63	0.53	0.24	0.26	0.43	0.52
Percentiles of Treatment Effect								
5th	0.00	-3.01	-0.82	-5.69	0.00	-0.01	-0.93	-4.92
25th	0.00	0.00	0.00	-2.67	0.00	-0.01	-0.17	-2.41
50th	0.00	0.00	0.17	0.21	0.00	-0.01	0.00	0.18
75th	2.04	2.44	0.59	3.13	0.00	0.54	0.39	2.63
95th	4.19	5.00	1.21	5.89	3.52	4.12	1.49	4.81
	Food Consumption				***Food Consumption***			
Average Treatment Effect	0.04				0.06			
	(-0.02,0.09)				(0.00,0.12)			
Impact Standard Deviation	0.30	2.12			0.38	2.08		
	(0.26,0.36)	(2.09,2.16)			(0.33,0.42)	(2.04,2.11)		
Proportion Positive	0.49	0.51			0.54	0.51		
Percentiles of Treatment Effect								
5th	-0.43	-3.49			-0.55	-3.35		
25th	-0.07	-1.68			-0.16	-1.61		
50th	0.00	0.03			0.03	0.05		
75th	0.14	1.69			0.27	1.69		
95th	0.54	3.42			0.65	3.33		

Notes: To generate these estimates treatment and control groups are first ranked from highest to lowest and then matched based on their rank. Under rank preservation (odd columns) those with the highest values in treatment are matched to those with the highest values in control. Under rank inversion (even columns) those with the highest values in treatment are matched to those with the lowest values in control. We then subtract the control value from treatment to get an estimate of the "individual treatment effect". We then report the average of this value across the sample as the average treatment effect, the standard deviation of these values, and the fraction of them that are positive. Finally, we sort these values from largest to smallest and report the percentiles of the sorted treatment effects. Panel A reports results from a measure of inverse hyperbolic sine (I.H.S.) transformation for bank savings and total savings, while Panel B reports the results from I.H.S transformation for food consumption. The Uganda and Malawi data are from Dupas et al. (2018). Bootstrapped confidence intervals with 1000 replications in parantheses.

Table A9: Heterogeneity in Microinsurance (IHS transformation)

	Gross Income				Consumption			
	Insurance		*Insurance + Grant*		*Insurance*		*Insurance + Grant*	
	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ATE	0.40		0.43		0.22		-0.02	
	(0.31,0.49)		(0.34,0.51)		(0.12,0.33)		(-0.14,0.10)	
Impact Standard Deviation	0.91	1.15	0.84	1.13	1.06	1.42	1.14	1.47
	(0.84,0.98)	(1.09,1.21)	(0.77,0.90)	(1.06,1.20)	(0.97,1.14)	(1.34,1.49)	(1.00,1.27)	(1.35,1.59)
Proportion Positive	0.69	0.67	0.82	0.72	0.64	0.61	0.49	0.48
Percentiles of Treatment Effect								
5th	-1.10	-1.14	-0.87	-1.28	-1.66	-2.14	-1.22	-2.71
25th	-0.14	-0.23	0.12	-0.10	-0.35	-0.70	-0.51	-0.74
50th	0.42	0.41	0.39	0.34	0.25	0.24	-0.02	-0.03
75th	0.95	1.00	0.75	1.17	0.72	1.13	0.42	0.97
95th	1.73	2.23	1.70	2.09	1.56	2.30	1.53	2.21

Notes: To generate these estimates treatment and control groups are first ranked from highest to lowest and then matched based on their rank. Under rank preservation (odd columns) those with the highest values in treatment are matched to those with the highest values in control. Under rank inversion (even columns) those with the highest values in treatment are matched to those with the lowest values in control. We then subtract the control value from treatment to get an estimate of the "individual treatment effect". We then report the average of this value across the sample as the average treatment effect, the standard deviation of these values, and the fraction of them that are positive. Finally, we sort these values from largest to smallest and report the percentiles of the sorted treatment effects. The left panel reports results from a measure of inverse hyperbolic sine (I.H.S.) transformation for gross income, while the right panel reports the results from a measure of I.H.S. transformation for consumption. The Gross Income and Consumption data are from the first year of intervention in Karlan et al. (2014). Bootstrapped confidence intervals with 1000 replications in parantheses.

Table A10: Variance Decomposition

	Egypt (Qena)		Bosnia		Egypt (Alexandria)		South Africa	
	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Firm Profits	2.6	3.4	1.6	6.0	11.5	14.0		
Consumption	2.4	3.3	1.2	8.0	2.0	4.0	5.1	9.4
	Uganda		**Malawi**		**Ghana: Insurance**		**Insurance + Grant**	
	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion	Rank Preservation	Rank Inversion
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Bank Savings	6.8	5.6	6.2	5.3				
Total Savings	16.1	12.4	5.5	3.8				
Consumption	4.4	6.3	4.6	3.0	8.4	11.7	2.5	9.6
Gross Income					12.3	33.6	29.0	37.5

Notes: This table reports the results from a variance decomposition exercise where we regress the estimated treatment effect on several baseline variables. We use the variables that were reported in the baseline balance tables from each paper. Each cell can be interpreted as the percent of the total variance in treatment effects that can be explained by those baseline covariates.