World Development 128 (2020) 104773

Contents lists available at ScienceDirect

World Development

journal homepage: www.elsevier.com/locate/worlddev

What do we know about the impact of microfinance? The problems of statistical power and precision ☆

Mahesh Dahal [a], Nathan Fiala [b,*]

[a] University of Connecticut, Agricultural and Resource Economics, United States
[b] University of Connecticut, Makerere University and RWI – Leibniz Institute for Economic Research, United States

ARTICLE INFO

Article history:
Accepted 22 November 2019

Keywords:
Microfinance
RCTs
Replication
Power calculations

ABSTRACT

We review all eight randomized control trial studies of microfinance published in peer-reviewed journals. The studies generally show no or minimal impact from providing microloans to clients and have led many researchers and policy makers to conclude that microfinance has been proven to have little or no positive impacts on people's lives. We review these studies in detail and find four main results. First, we are able to replicate the results using the researcher's original data. Second, we observe that while the results are generally insignificant at traditional levels, most estimated coefficients are large. Third, every one of the studies is underpowered to detect reasonable effect sizes, often due to low take-up of the financial product offered. Pooling the data from the studies together improves power for most outcomes, but minimum detectable effect sizes are still large. Finally, when we run analysis on a pooled sample, we find a treatment effect on business profits, business revenue and household assets, significant at the 1% level. We argue that existing research on the impact of microfinance is generally underpowered to identify impacts reliably and suggests that we still know very little about the impact of microfinance. We end by discussing ways to improve future research.

1. Introduction

Microfinance has a long and complicated history with policy makers and researchers. Many initial proponents argued that lack of access to formal finance was a critical part of why people remained poor in developing countries. Despite initial quasi-experimental studies on the impact of microcredit showing large promise in poverty reduction, mainly among female borrowers (Pitt & Khandker, 1998), recent experimental studies of microfinance have shown the impact of access to microcredit is not as transformative as it was once thought to be. The large promises of microfinance solving world poverty were not panning out.

Decades after the beginning of the microfinance movement, there is still little conclusive evidence on the impact of microfinance on the lives of the poor. Any credible attempt at identifying the impact of microcredit on the wellbeing of people needs to overcome the concerns of double selection in credit markets—lenders selecting potential borrowers as well as borrowers self-selecting to borrow. Impact evaluations based on randomized control trials (RCT) have been increasingly used in the past decade to overcome identification issues. However, as we discuss in detail below, findings from RCTs remain inconclusive, mostly owing to low net take-up[1] of microfinance products that presents a statistical power challenge for RCT studies of microfinance.

Karlan and Zinman (2011) was one of the first studies to look at the impact of microfinance on poor households using data from an RCT in the Philippines. They found that the microcredit intervention they studied did not lead to bigger businesses, higher income, or higher subjective well-being, but instead resulted in better risk management, fewer businesses, and lower subjective well-being among those who received the microcredit treatment.

Several more recent randomized studies also look at the impact of the traditional microfinance model. Six randomized studies published in a 2015 special issue of the *American Economic Journal: Applied Economics* have come to a similar conclusion, showing lack

☆ We thank Chris Blattman, David McKenzie, Rachael Meager and Jörg Peters for their helpful comments. Funding for research assistant time for this study was provided by Opportunity International. The results and opinions are those of the authors only.

* Corresponding author.
E-mail addresses: mahesh.dahal@uconn.edu (M. Dahal), nathan.fiala@uconn.edu (N. Fiala).

[1] Net take-up rate refers to take-up rate differential among treatment and control groups. In this paper, we use net take-up rates and net compliance rates interchangeably. Take-up rate refers to the take-up of the treatment product among those who were assigned into treatment group.

https://doi.org/10.1016/j.worlddev.2019.104773
0305-750X/

of the "transformative" role of microfinance on the lives of poor households. More recently Fiala (2018) found positive impacts for microfinance, but only for male borrowers and only in the short run.

We closely review here the evidence presented in these eight experimental studies on the impact of increased access to microfinance. Contrary to recent work that has simply taken these designs at face value (for instance in Meager, 2019), we look carefully at the designs of each study and their ability to answer the questions they pose. We replicate the results of the studies using data provided by the authors[2], conduct ex-post power calculations to determine what can be reasonably expected from each study, pool the data to run a better powered test, and discuss different contexts and models these studies evaluate. Table 1, adopted from Banerjee, Karlan, and Zinman (2015), provides information on key features and a summary of each of the studies.

Table 1 also presents the rates of access to credit among the study population during the baseline of seven of the eight studies. Five of these study areas have credit access rate of 50 percent or more. This relatively high credit access rate poses significant obstacles to attempts of finding average treatment effects of microfinance interventions, both because those with outstanding loans are less likely to take-up microfinance, and the treatment effects are not likely to be the same for those with and without prior access to credit.[3]

As expected, we are able to replicate all of the findings in the eight individual studies. We find that the coefficients for many outcomes are large when compared to control group means, but most are not statistically significant at traditional levels. We find evidence that this is likely due to serious power issues in each of the studies. The results of ex-post power calculations for intent to treat (ITT) estimates for the impact of microfinance on several profit and income related outcomes for the individual studies shows that most ITT estimates are significantly under-powered. The minimum detectable effect sizes (MDES) for main outcomes are as high as 0.4 standard deviations or up to 230% of control mean under perfect compliance, and as high as 1.54 standard deviations or up to 1,000% of control mean under actual compliance. Median MDES under perfect compliance is 0.16 standard deviations (22% of control mean) while it is 0.73 standard deviations (132% of control mean) under actual compliance.[4]

We then combine the data into one sample and run power calculations on the pooled data. Even though the studies were designed independently, endline measures of outcomes are mostly comparable across the studies, and so most outcomes can be easily combined. We find that this improves MDES to between 0.04 and 0.12 standard deviations (8% and 115% of mean of outcome for control group) under perfect compliance for most main outcomes. Using actual compliance rates, we find MDES between 0.17 and 0.48 standard deviations (32% to 458% of control mean). We conclude that the individual studies are significantly underpowered to make inferences on the outcomes they focus on. When combining data, the situation is significantly improved, but is still not ideal.

Because pooling data significantly improves power, we conduct analysis on the full sample, representing over 33,000 participants, running a single OLS regression. Our analysis follows Banerjee et al. (2015b) by assuming a common slope. We include country and region[5] fixed effects and introduce household controls[6] when available to further improve power, weight the samples to account for cross-study imbalance in sample sizes[7], and adjust for timing of the surveys and purchasing power parity differences. We find impacts on business profits of at least 27.4% above the control group, significant at the 1% level. Using our preferred specification that includes household controls and regional dummies we find 40.3% (0.05 standard deviations) impacts on business profits, significant at 1% level. We do not find statistically significant impacts on total consumption, but a significant increase of household assets.

As we present tests of several individual hypotheses, there is a potentially high probability of rejecting a true null hypothesis, just by chance. In order to rule out this possibility we also present p-values adjusted for multiple-hypothesis testing. We follow Romano and Wolf (2005) who develop a stepwise multiple-testing procedure that asymptotically controls the family-wise error rate. We find that impact estimates for profits and assets are statistically significant at standard levels even after accounting for multiple hypothesis testing.

Our results from fully pooling the data are similar to a recent study by Meager (2019), who looks at the data from seven randomized experiments, including Karlan and Zinman (2011) and the six *AEJ: Applied Economics* studies. She performs both a Bayesian hierarchical analysis with partial pooling and an OLS analysis on a fully pooled sample. She does not weight the samples and finds a coefficient estimate of 7.3[8]. This is very close to our unweighted coefficient and represents a 20% treatment effect size. However, Meager (2019) argues that fully pooling is not appropriate as there may be heterogeneity in the samples. She finds a 7% effect size in the hierarchical analysis model and argues that the average effect of microcredit is positive but small relative to control group average levels, with a reasonably high chance of effectively zero impact.

While fully pooling the sample has limitations, Meager (2019) finds that heterogeneity was only moderate for these studies.[9] This means that fully pooling the sample is not in fact as problematic as initially thought. We see the issue of power as being of more immediate concern, and partial pooling of the data does not alleviate the power issues we highlight in the individual studies. By fully pooling the data, weighting the studies equally and controlling for region fixed effects as well as controlling for household level covariates, we are able to improve power and control for potential bias from heterogeneity across the studies. The *meta*-analysis literature has argued that moderately large heterogeneity among the studies' tar-

[2] Note that our replications are simply push button. Bédécarrats, Guérin, Morvant-Roux and Roubaud (2015) look closer at the data in Crépon, Devoto, Duflo, and Pariente (2015) and find significant errors that call into question the original analysis.

[3] As Wydick (2016) and Ogden (2016) point out, evaluating microfinance in areas where access to credit is high also means these studies are capturing marginal borrowers. This population likely has lower returns to finance and so offers an additional challenge when trying to determine if finance in general helps people.

[4] We focus on estimating the intention to treat effects as this is what was done in the original studies. We do not estimate local average treatment effects (LATE) as this simply increases the size of the estimated coefficients and does not affect power or statistical significance. We discuss extensively the compliance rate for studies in this paper only to highlight power issues.

[5] Regions are selected based on the study design and does not necessarily refers to geographic regions for these countries. Regions include: 14 bank branches in Bosnia, 16 Woredas in Ethiopia, a single block in India, 45 super clusters in Mexico, 5 provinces in Mongolia, 81 village pairs in Morocco, 8 branches in the Philippines and 4 districts in Uganda.

[6] Household controls include gender, education, and age of the household head or the survey respondent for countries for which information on household head is not collected. Household controls not available for the Mexico data.

[7] We conduct both unweighted and weighted (our preferred approach) analysis, which allows us to control for the different sample sizes across studies. By weighting we are able to treat each study as equal to all others. We believe this is an important adjustment for the samples as some country studies, such as Mexico, represents almost half of the total sample, while India and Morocco are about 1/6 of the sample and Bosnia, where compliance with treatment was best, is less than 3% of the total sample.

[8] We replicate the results of Meager (2019) in Appendix Table A5 for comparison and find the same results as she does.

[9] Meager (2019) finds that for seven of the studies included in this paper, on average, 60% of the observed variation in treatment effects can be attributed to sampling variation across studies and that the genuine underlying heterogeneity is moderate and smaller than previously thought.

Table 1
Loan information and sampling for the eight studies.

	Bosnia and Herzegovina	Ethiopia	India	Mexico	Mongolia	Morocco	Philippines	Uganda
Unit of randomization	1,196 individual applicants	133 peasant associations	104 neighborhoods	238 clusters (neighborhoods or villages)	40 villages	162 villages	Individual applicants	Individual applicants
Gender of borrowers	41%	13% female household head	100%	100%	100%	7% female household head)	85% female	61.5% female
Targeted to Microentrepreneurs?	Yes (91 percent of responders planned to invest in new or existing business)	Yes (Plans for starting business considered "salient" criteria	No	Yes (Has business or interested in starting one)	Yes	Yes	Yes	Yes
Sampling frame	Marginal loan applicant considered too risky and "unreliable" to be offered credit as regular borrowers under the terms above	Random selection of households	Households with at least one woman age 18–55 that have resided in the same area for at least three years	Mexican women ages 18–60 who either have a business/ economic activity, would start one if they had enough money, or would consider taking credit from an institution	Women who met eligibility criteria and signed up to declare interest in receiving loan from lender	(1) Households deemed likely borrowers, (2) random selection of households	Marginal borrowers. Sample drawn from a universe of applicants who at eight of the lender's nine branches between 10 February 2006 and 16 November 2007	Enterprises contacted from a census of businesses operating in four districts in Uganda. The final sample composed of the 1550 individuals who twice expressed interest in trainings and loans
Loan term length	Average 14 months	12 months	12 months	4 months	3–12 months group (average 6 months); 2–24 months individual (average 8 months)	3–18 months (average 16 months)	13 weeks	12 months
Repayment frequency	Monthly	Borrowers were expected to make regular deposits and repayments	Weekly	Weekly	Monthly	Weekly, twice monthly, or monthly	Weekly	Monthly
Interest rate	22 percent APR	12 percent APR	24 percent APR (12 percent nondeclining)	110 percent APR	26.8 percent APR	14.5 percent APR	60% APR	20% APR
Market interest rate	27.3 percent APR	24.7 percent APR	15.9 percent APR	145.0 percent APR	42.5 percent APR	46.3 percent APR	60% APR	26% APR
Liability	Individual lending	Group (joint liability)	Group (joint liability)	Group (joint liability)	Two treatment arms: group (joint liability) and individual	Group (joint liability	Individual lending	Individual lending
Baseline credit access rate	58.3%	13.1%	68%	53.7%	57.3%	24% (including 16% from utility companies and 6% informal)		49%
Sample size	994	6,263 (endline)	6,811	~16,150	611	4,934	1,113	765
Net compliance rate (Any MFI loan)	43.9 ppts	25.2 ppts	8.4 ppts	6.9 ppts	37 ppts (approx.)	9.0 ppts (approx.)	71 ppts	23 ppts
(Any loan)	19.3 ppts	25.2 ppts	0 (approx.)	5.1 ppts	25.7 ppts	7.6 ppts		23 ppts

Source: Based on Banerjee et al. (2015) and original studies.

get populations, intervention content and modalities, and other aspects may be addressed using study-level along with participant-level covariates (Bangdiwala, 2016; Morton, Adams, Suttorp, & Shekelle, 2004).[10]

It is also important to note that the size of coefficients in each of the original studies is often quite large, even when not statistically significant at traditional, frequentist metrics. If one ignores statistical significance testing, as has been advocated by some, the effects from each of the studies is quite large.

In summary, we differ from Meager (2019) in several ways. First, we do not take the designs and data at face value but instead offer a critical perspective on each study. Second, we directly address one of the biggest issues in studies of microfinance: statistical power. Third, we conduct weighted analysis to ensure any one study, such as Mexico, does not dominate the sample. Finally, we include all of the relevant published studies on this topic in our analysis.

Our work in this paper builds off of others who have looked closely at the studies we discuss here. To the best of our knowledge, McKenzie (2012) was the first to note low statistical power in one of the eight studies we discuss here. He finds that a sample size of 15 million would have been needed to obtain the power to identify effect sizes of 10% in the India study (Banerjee et al., 2015a).

Our main contribution here is to present a clear discussion of power issues in these studies. While the original authors of the experiments acknowledge there are issues with statistical power in these studies, we present evidence of the magnitude of the problem and discuss what this means for our ability to interpret the results. Banerjee et al. (2015), in the introductory paper to the special *AEJ: Applied Economics* issue, note that low take-up rates of credit presents a statistical power challenge to most randomized impact evaluations of microfinance published in the special issue. However, they never quantify just how big of an issue this is.

Based on our review of the evidence presented in the eight experiments, we conclude that many of the studies that are presented by some as definitive evidence of the (lack of) effect of microfinance in fact fail to provide definitive answers on the impact of microfinance, at least individually. But this does not mean that our estimates of impacts from the pooled sample should be taken as the true impact of microfinance. While the most likely reason for our finding impacts when the individual studies did not is improved power, there are still some power issues in the pooled sample. As Gelman and Carlin (2014) argue, in under-powered studies, statistically significant results can still be misleading. In some cases, it is possible to overstate the magnitude of effects, and even get the sign wrong.[11]

The problem of statistical power is pervasive in empirical studies. McKenzie and Woodruff (2013) show significant power issues in all of the 12 experimental studies of business skills training programs they review. Ioannidis, Stanley and Doucouliagos (2017) review 6,700 empirical economics studies and find more than half of them are under-powered. While the issues of low power can be a problem for many studies, not just microfinance, given the scope of the microfinance sector and debates around it among researchers and policy makers, we believe it is important to understand the size of the problem.

Unlike the perception among many critics of microfinance, the studies reviewed here do not discredit the role of microcredit in poverty alleviation and improving livelihoods of poor households. Nor does combining the samples together definitively show impacts. What these results do suggest is that the impact of the microfinance programs studied in these experiments is not well known.

One could interpret the original results to say that microfinance is probably not as transformative as it was once believed. But this is often used as a straw man by critics. The initial, unusually high expectations placed on microfinance to tackle mass poverty and fuel sustained economic growth should not be the basis for dismissing potentially more modest findings of the impact of microfinance on improving livelihoods of poor households. Many of the null results found in the original eight studies include economically meaningful effect sizes but could not be taken as conclusive due to power issues.

This paper proceeds as follows. In section 2 we present the results from our replication of the eight individual studies and from pooling the data. In section 3 we present our discussion of statistical power. In section 4 we offer our interpretation of what can reasonably be concluded from existing work on microfinance, including other studies that have been produced more recently. We end in section 5 with a discussion of what we feel is missing from the research on microfinance and what can realistically be done to design studies with high chances of providing sufficient power to detect the impact of microfinance.

2. Replication of previous experiments

We present in this section our replication of the main results for the eight experimental studies. In addition to conducting a pure replication, we also pool data for common variables in all eight studies and run regressions for common outcome variables on the combined dataset. Since there are large differences in sample sizes for the different studies, we report estimates from weighted regressions where each country study has an equal weight. We also provide estimates from unweighted regressions in the appendix. We begin by discussing the studies themselves.

2.1. The eight experiments

The eight studies we discuss here include experiments in Bosnia and Herzegovina (Augsburg et al., 2015), Ethiopia (Tarozzi, Desai, & Johnson, 2015), India (Banerjee et al., 2015), Mexico (Angelucci, Karlan, & Zinman, 2015), Mongolia (Attanasio, Augsburg, & De Haas, 2015), Morocco (Crépon, Devoto, Duflo, & Parienté, 2015), the Philippines (Karlan & Zinman, 2011), and Uganda (Fiala, 2018). We obtained the data and analysis code for seven of these studies online, downloaded on May 2017. Data and code for Tarozzi et al. (2015) was kindly made available by the authors upon request.

Five of the eight microfinance studies reviewed here measure the impact of group liability loans. Attanasio et al. (2015) tested the impact of both group liability and individual liability loans in Mongolia, but they focus their analysis on group liability lending. Augsburg et al. (2015), Karlan and Zinman (2011), and Fiala (2018) measure impact of individual lending in Bosnia and Herzegovina, the Philippines, and Uganda, respectively. Three of the studies were conducted in rural settings, three in urban settings, one in a semi-urban setting and one in both urban and rural setting. Three studies provided microcredit to women only and five provided to both men and women. Interest rates (APR) varies widely between 12% in Ethiopia to 110% in Mexico.

While the studies in Bosnia, the Philippines, and Uganda randomized loan access at the individual level, the remaining studies randomized microcredit programs at village, community or neigh-

[10] We thus essentially adopt a fixed-effects meta-analysis approach. This approach assumes that there is little heterogeneity in treatment effects across the studies, which is reasonable given the results of Meager (2019) that the level of heterogeneity in treatment effect was only moderate for these studies.

[11] It should be noted that the power issues we discuss here apply equally to studies with a large number of participants or just a few. Whatever the sample size, if compliance in a study is very low, as is the case in most microfinance studies, the probability of any type of error increases significantly.

borhood level. In India, 52 out of 104 urban neighborhoods were randomly selected to be served by a microfinance institute. Tarozzi et al. (2015) uses data from a microcredit and family planning program in Ethiopia that was conducted using a 2 × 2 factorial design where 133 local administrative units (PAs) were randomly assigned to one of four groups: microlending only, family planning program only, both, or none. As the main focus of the paper is the impact of microfinance, we limit the sample to microfinance only treatment groups and the pure control groups. In Mexico, 238 geographic clusters that were not served by the partner microfinance institution (MFI) were randomized to receive credit access and loan promotion program. In Mongolia, 40 villages were randomized to either receive group liability (15), individual liability (15), and no (control group) microcredit program (10). In Morocco, the evaluation was implemented in 162 villages, divided into 81 pairs of similar villages. The pairs were chosen in areas the partner MFI was planning to start its operation. One village in each pair was randomly selected for the partner MFI to start its lending program.

All eight studies measure the impact of increased access to microfinance. Most of these studies conduct an intent-to-treat (ITT) analysis, with a few focusing on a sample of "likely borrowers." In most cases, the average effect of easier access to microfinance on those who are primary targets of microfinance institutions is evaluated.

It could be the case that the impact of microfinance is greatest for those who are truly credit constrained. If this is the case, then it is not surprising that the studies reviewed here do not find much impact. Five of the eight studies were conducted in contexts where at least 40 percent of the households already had access to microfinance, and the interest rates charged by partner microfinance institutes of these studies were not much better than the market interest rates. We again refer to Table 1 for the summary of the eight studies. 58 percent of respondents in Bosnia, 68 percent in India, 54 percent in Mexico, and 57 percent in Mongolia, and 44 percent in Uganda had at least one outstanding loan at baseline. 36 percent of control group respondents in the Philippines had at least one outstanding loan. Two out of eight studies (Ethiopia and Morocco) were conducted in settings where households seem to have very limited access to microcredit. 13 percent respondents in Ethiopia and less than 25 percent (including 16 percent of households with loans from utility companies) in Morocco reported to have at least one outstanding loan from any source. Net compliance rates for loans from an MFI or a bank are generally very low and vary significantly across countries. Difference in percentage of those with loans in treatment and control groups are 49.3 percentage points (ppts) for Bosnia, 25.2 ppts for Ethiopia, 8.4 ppts for India, 11.5 ppts for Mexico, 25.7 ppts for Mongolia, 16.7 ppts for Morocco 9.4 ppts for the Philippines, and 10 ppts for Uganda. This net compliance issue will be the main driving force behind the issues with power that we discuss below.

2.2. Results from the individual studies

We now present the results of replicating the individual studies. Because we find broadly the same results as the original authors, we do not present them in a table, but instead describe some of the more important findings.

The first impact we review is whether treatment led to an increase in access to finance. All of the treatments led to increased access to loans from microfinance institutes (MFIs) as compared to control groups. However, access to any type of loan did not significantly increase in treatment areas for some studies. Given that the interest rates provided by the MFIs are very similar to the market interest rates, it is important to observe increase in credit from all sources in order to expect any impact of the microcredit intervention. Net take-up of MFI loans varied significantly across studies, with treatment groups in Mexico (6.9), India (8.4), Morocco (9.0), the Philippines (9.4), and Uganda (10) only slightly more likely to have an MFI loan than control group. Ethiopia (25.2), Mongolia (37) and Bosnia (43.9) saw greater access to any MFI loans. Access to any type of loan among the treatment group was higher by only 5.1 ppts in Mexico,7.6 ppts in Morocco, 10 ppts in Uganda, 19.3 ppts in Bosnia, 25.2 ppts in Ethiopia and 25.7 ppts Mongolia. In India, households in treated areas were not any more likely to have a loan from any source as compared to households in control areas.

There is some impact of increased access to microfinance on starting a business. Self-employment activity increased in Mongolia. Investment in business increased in India, Mexico, and Morocco. In India, 15 to 18 months after gaining access to loans, households are no more likely to have at least one business, but they invest more in the businesses they do have or the ones they start. There is an increase in the average profits of the businesses that were already in existence before microcredit intervention. In Mexico, households in treatment areas grow their businesses (both revenues and expenses increase), but there were no corresponding effects on business profits and entry. In Morocco, among those who take up microcredit loans, there are proportionally large average impacts on self-employment investments, sales and profits, although there is great heterogeneity in these effects. Although the impacts on profit are statistically indistinguishable from zero for most countries, the effect sizes are generally large and economically meaningful. For profit outcome, effect sizes as percentage of control mean are 23% in Bosnia, 68% in Ethiopia, 48% in India, 18% in Mongolia, 22% in Morocco, close to 0% in Mexico, 4.2% in Philippines, and 5.6% in Uganda.

None of the studies find impact on income. In Mongolia, there is positive impact of microcredit on total consumption as well as in food consumption. However, there is no impact on total consumption as well as food consumption in any other countries. Many of the null intent-to-treat results have confidence intervals that include economically meaningful effect sizes, particularly if one were to scale up the intent-to-treat estimates to infer treatment-on-the-treated effects.

2.3. Results for combining the studies

We next discuss the results of pooling the individual studies into a single dataset. We include country and region[12] fixed effects in the analysis. In our preferred specification we add region fixed effects as well as include household controls. As household controls are not available for the Mexico study, regressions with household controls is run on data for seven countries only. We are only able to control for household head or respondent's gender, age, and education. We control for regional effects even though the original studies do not as this is a more common approach to analysis of RCTs today as it increases statistical power without loss of internal validity. Our results are generally robust to including this control or not.

We obtain significant improvement in the precision of the estimates in the combined dataset. We find large and highly significant treatment effects on assets, profits, and revenues. However, we do not find any impact on durable consumption, food consumption and total consumption. We also do not find any impact on wage income.

As we discuss in the next section, some of the outcomes measured are severely underpowered, even when combining data.

[12] Regions are selected based on the study design and does not necessarily refers to geographic regions for these countries. Regions include: 14 bank branches in Bosnia, 16 Woredas in Ethiopia, a single block in India, 45 super clusters in Mexico, 5 provinces in Mongolia, 81 village pairs in Morocco, 8 branches in the Philippines and 4 districts in Uganda.

We are not able to control for baseline covariates, as they differ for each study, which is also a problem for proper inference of results. However, we are able to include region fixed effects for 7 out of 8 studies, which does help improve the precision of some of the estimates.

Results of the estimation using the combined sample is presented in Tables 2–4. Sample size for different outcome variables differ as some studies are missing variables common to other studies. For most of the common variables, we have a sample size of about 33,000 observations for all eight countries and about 17,200 observations for sample that excludes Mexico. Money figures such as loan amount, income, consumption values are expressed in purchasing power parity (PPP) 2009 USD terms for all countries. Asset values are standardized to standard deviation of one and mean zero for all countries. Following the schedule in most of the papers, all income, profit, and consumption variables are expressed in fortnightly terms. Estimates presented in Tables 2–4 are weighted to ensure equal weights are placed for each study. Unweighted estimates are presented in the Appendix, Tables A1–A3.

We present tests of several individual hypotheses which implies a potentially large probability of rejecting a true null hypothesis of no impact by mere chance. In addition to robust standard errors, we present p-values adjusted for multiple-hypothesis testing in square brackets for each outcome tested. We obtain adjusted p-values following a stepwise multiple-testing procedure that asymptotically controls the family-wise error rate following Romano and Wolf (2005). Our results from combined dataset on outcomes for loan-take up as well as profits and assets hold even after correcting for multiple-hypothesis testing.

Table 2 presents combined results on access to credit. Combined results suggest that microfinance treatment results in 12 ppts (5.9 ppts unweighted) increase in access to any type of loan, 21 ppts (14.5 ppts unweighted) increase in access to program MFI loans among persons/households in treatment areas. Combined results also suggest significant increase in loan amount among treatment group, mostly resulting from increase in loan amount from program MFIs.

Table 3 presents combined results on the impact of microfinance on wage income and profits from self-employment activities. We find a significant effect of microfinance intervention on total business profits (40% of control mean or 0.05 standard deviations) and production/revenues (19% of control mean or 0.05 standard deviations). We also find significant impacts on assets accumulation. We do not find any significant effect on business expenses and wage income. Although statistically insignificant, the coefficient on wage income is negative, suggesting that there may be crowding out of income sources as a result of increases in business profits.

Combined results on the impact of microfinance on household consumption are presented in Table 4. We do not find significant impacts on total consumption, food consumption durable goods or temptation goods.

2.4. Interpretation by the original authors

Do the modest net take-up rates of microcredit in these studies suggest anything about the effectiveness of microfinance in helping microentrepreneurs grow their businesses and improve consumption? Banerjee et al. (2015) suggest that the low-take up rates in these studies is the prima-facie case against the notion of microfinance being a panacea for poverty alleviation. Microfinance may not be a panacea for poverty, but we still do not know much about the relative effectiveness of microfinance compared to other development tools for improving livelihoods of the poor. The

Table 2
Loan take-up for the combined sample (Weighted).

	(1) Took loan offered	(2)	(3)	(4)	(5) Has any loan	(6)	(7)	(8)	(9) Amount of loan taken	(10)	(11)	(12)	(13) Amount of any loan	(14)	(15)	(16)
Assigned to treatment	0.31***	0.32***	0.21***	0.62***	0.12***	0.12***	0.13***	0.25***	234.8***	240.8***	284.0***	0.56***	371.8***	397.0***	429.4***	0.05***
	(0.01)	(0.01)	(0.02)	(0.06)	(0.02)	(0.01)	(0.01)	(0.03)	(26.6)	(27.2)	(31.3)	(0.06)	(93.6)	(101.4)	(115.9)	(0.02)
	[0.00]	[0.00]	[0.00]	-	[0.00]	[0.00]	[0.00]	-	[0.00]	[0.00]	[0.00]	-	[0.00]	[0.00]	[0.00]	-
Country FE	Yes	-	-	-	Yes	-	-	-	Yes	-	-	-	Yes	-	-	-
Region FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Household controls	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
Control mean	0.062	0.062	0.062	-	0.509	0.509	0.509	-	121.716	121.716	121.716	-	1,412.385	1,412.385	1,412.385	-
N	34,536	34,536	17,045	17,045	34,387	34,387	18,089	18,089	33,984	33,984	17,045	17,045	34,349	34,349	18,089	18,089
Adjusted R2	0.27	0.29	0.18	0.18	0.18	0.23	0.27	0.27	0.13	0.14	0.11	0.11	0.07	0.07	0.07	0.07
F-Statistics	481.6	554.6	27.3	27.3	61.9	93.2	23.2	23.2	78.0	78.6	21.3	21.3	15.8	15.3	8.6	8.6
Treatment effect as % control mean	511.3%	527.0%	343.9%	-	23.1%	22.8%	24.7%	-	192.9%	197.8%	233.3%	-	26.3%	28.1%	30.4%	-

Notes: Significance levels: (***1%, **5%, *10%). Standard errors in brackets clustered as in original studies. P-values adjusted for multiple hypothesis testing are presented in square brackets. Significance stars based on unadjusted p-values. Regions include: 1 in India, 14 branches in Bosnia, 16 Woredas in Ethiopia, 45 super clusters in Mexico, 5 provinces in Mongolia, and 81 village pairs in Morocco, 8 branches in the Philippines and 4 districts in Uganda. All monetary variables are expressed in 2009 PPP USD terms using data obtained from WDI. Observations are weighted so that the sum of weights is equal for all countries. Household controls include age, education, and gender for household head or respondent. Dependent variable for the fourth regressions for each outcome (columns 4, 8, 12, and 16) are standardized to mean zero and standard deviation of 1.

Table 3
Income outcomes for the combined sample (Weighted).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	Assets				**Profits (2 weeks)**				**Revenue (2 weeks)**				**Wage income (2 weeks)**			
Assigned to treatment	0.03*	0.04*	0.05*	0.05*	15.5***	18.7***	22.8***	0.05***	20.9	25.3	46.7*	0.05*	−96.2	−125.0	−141.9	−0.15
	(0.020)	(0.018)	(0.023)	(0.023)	(6.454)	(6.692)	(8.351)	(0.019)	(22.995)	(24.533)	(26.498)	(0.030)	(106.5)	(111.6)	(127.1)	(0.138)
	[0.355]	[0.060]	[0.060]	–	[.]	[0.016]	[0.020]	–	[0.669]	[0.566]	[0.151]	–	[0.633]	[0.566]	[0.335]	–
Country FE	Yes	–	–	–	Yes	–	–	–	Yes	–	–	–	Yes	–	–	–
Region FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
HH controls	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
Control mean	−0.075	−0.075	−0.075	–	56.518	56.518	56.518	–	246.872	246.872	246.872	–	387.046	387.046	387.046	–
N	32,116	32,116	13,929	13,929	34,872	34,872	17,371	17,371	33,341	33,341	17,136	17,136	33,224	33,224	17,025	17,025
Adjusted R2	0.02	0.04	0.06	0.06	0.08	0.09	0.09	0.09	0.16	0.17	0.16	0.16	0.02	0.03	0.03	0.03
F-Statistics	2.8	5.5	12.8	12.8	5.8	7.8	4.8	4.8	0.8	1.1	6.1	6.1	0.8	1.3	1.0	1.0
Treatment effect as % control mean	−44.7%	−56.3%	−72.2%	–	27.4%	33.2%	40.3%	–	8.5%	10.2%	18.9%	–	−24.9%	−32.3%	−36.7%	–

Notes: Significance levels: (***1%, **5%, *10%). Standard errors in brackets clustered as in original studies. P-values adjusted for multiple hypothesis testing are presented in square brackets. Significance stars based on unadjusted p-values. Regions include: 1 in India, 14 branches in Bosnia, 16 Woredas in Ethiopia, 45 super clusters in Mexico, 5 provinces in Mongolia, and 81 village pairs in Morocco, 8 branches in the Philippines and 4 districts in Uganda. All monetary variables are expressed in 2009 PPP USD terms using data obtained from WDI. Assets values are standardized to have mean zero and standard deviation of 1. Observations are weighted so that the sum of weights is equal for all countries. Household controls include age, education, and gender for household head or respondent. Dependent variable for the fourth regressions for each outcome (columns 4, 8, 12, and 16) are standardized to mean zero and standard deviation of 1.

Table 4
Consumption outcomes for the combined sample (Weighted).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	Total consumption (2 weeks)				Food consumption (2 weeks)				Durable consumption (2 weeks)				Temptation goods (2 weeks)			
Assigned to treatment	18.1	20.4	24.6	0.09	3.0	3.0	3.8	0.03	5.2	5.6	5.4	0.05	−1.1	−1.3	−1.4	−0.05
	(19.9)	(19.1)	(23.6)	(0.09)	(4.1)	(4.2)	(5.6)	(0.05)	(4.5)	(4.1)	(4.1)	(0.04)	(0.87)	(0.88)	(1.03)	(0.04)
	[0.57]	[0.40]	[0.47]	–	[0.57]	[0.43]	[0.47]	–	[0.57]	[0.39]	[0.47]	–	[0.28]	[0.37]	[0.47]	–
Country FE	Yes	–	–	–	Yes	–	–	–	Yes	–	–	–	Yes	–	–	–
Region FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
HH controls	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
Control mean	275.832	275.832	275.832	–	150.528	150.528	150.528	–	45.217	45.217	45.217	–	21.997	21.997	21.997	–
N	15,377	15,377	13,918	13,918	32,885	32,885	14,944	14,944	13,899	13,899	13,872	13,872	30,380	30,380	13,918	13,918
Adjusted R2	0.17	0.21	0.08	0.08	0.24	0.26	0.28	0.28	0.06	0.07	0.07	0.07	0.07	0.08	0.10	0.10
F-Statistics	0.8	1.1	14.6	14.6	0.5	0.5	5.0	5.0	1.3	1.8	4.3	4.3	1.6	2.3	13.1	13.1
Treatment effect as % control mean	6.6%	7.4%	8.9%	–	2.0%	2.0%	2.5%	–	11.5%	12.3%	12.0%	–	−5.0%	−6.1%	−6.4%	–

Notes: Significance levels: (***1%, **5%, *10%). Standard errors in brackets clustered as in original studies. P-values adjusted for multiple hypothesis testing are presented in square brackets. Significance stars based on unadjusted p-values. Regions include: 1 in India, 14 branches in Bosnia, 16 Woredas in Ethiopia, 45 super clusters in Mexico, 5 provinces in Mongolia, and 81 village pairs in Morocco, 8 branches in the Philippines and 4 districts in Uganda. All monetary variables are expressed in 2009 PPP USD terms using data obtained from WDI. Observations are weighted so that the sum of weights is equal for all countries. Household controls include age, education, and gender for household head or respondent. Dependent variable for the fourth regressions for each outcome (columns 4, 8, 12, and 16) are standardized to mean zero and standard deviation of 1.

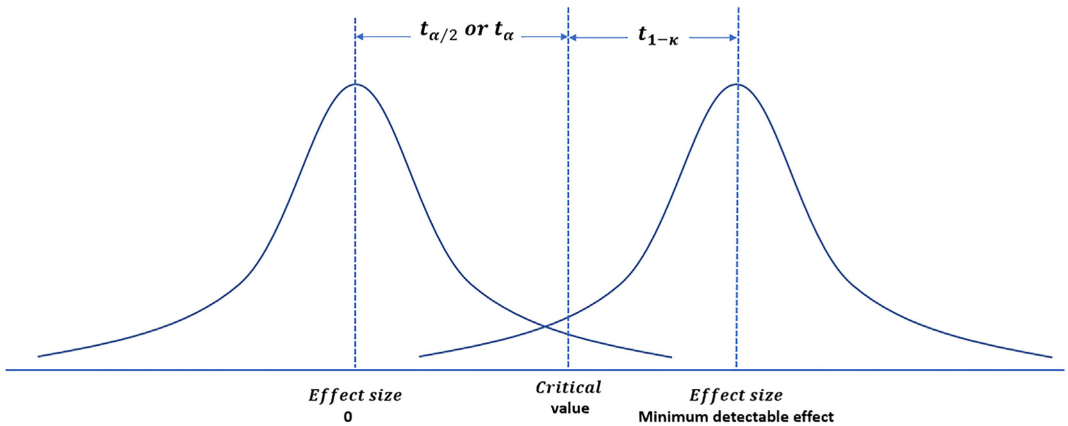

Fig. 1. Illustration of minimum detectable effect size multiplier.

eight studies considered here in this analysis are not able to show evidence of transformative effects of microfinance on the average borrower. However, the authors also caution the readers that the lack of transformative effects should not obscure other more modest but potentially important effects.

Summarizing the six papers, Banerjee et al. (2015) come up with the following conclusion: First, the studies do not find clear evidence of transformative effects of reductions in poverty or improvements in living standards. Second, the lack of transformative effect does not mean absence of modest but important effects on investment in business growth. There is convincing evidence that businesses expand, and some evidence that profits increase. Failure of business expansion to translate into improvements in living standards may be because of tradeoff between business and wage income and heterogeneity on profitability (larger firms more profitable than smaller). Third, the lack of transformative effects should not obscure other more modest, but potentially important, effects leading to increase in freedom of choice: improvements in occupational and consumption choice, female decision power, and improved risk management. Fourth, the studies find little evidence of negative effects even in the context of high interest rates. Fifth, the presence of heterogenous effects suggest that the impact of microfinance can be transformative for some. And finally, many of the null results are estimated imprecisely even when these effect sizes are economically meaningful.

Most importantly, Banerjee et al. (2015) believe that pooling across the studies would yield significant increases in business outcomes. Our results confirm this belief. While 40% effect sizes may not be transformative, they are still quite large relative to studies of other development programs. Our results suggest that microfinance can have important effects on business development and general economic growth. However, as we discuss in the next section, we must be cautious about the pooled sample impacts as even the pooled sample is still under-powered.

3. Power calculations for the eight studies

3.1. How to calculate power

Due to the low net take-up of the interventions tested, it is highly likely that there are serious statistical power issues with these studies. We next compute ex-post power calculations and minimum detectable effect sizes for the main outcome variables (credit access, self-employment, business profits, income, and consumption) for each of the eight studies.

Fig. 1 illustrates that the minimum detectable effect of an OLS estimator of impact is a multiple of the standard error of the impact estimator. The bell-shaped curve on the left represents a $t-$ distribution for a null hypothesis of no impact. For a positive impact estimate to be statistically significant at the $\alpha/2$ level with a two-tailed test, the estimate must fall to the right of the critical t-value, $t_{\alpha/2}$, of the distribution under null hypothesis of no impact.[13] The bell-shaped curve on the right represents a t-distribution for an alternative hypothesis that the true impact equals a specific minimum detectable effect. To have a probability $(1-\kappa)$ of detecting the minimum detectable effect it must lie a distance of $t_{1-\kappa}$ to the right of the critical t-value for the null hypothesis, where the probability $(1-\kappa)$ represents the level of statistical power. Hence the minimum detectable effect must lie a total distance of $\left(t_{\frac{\alpha}{2}}+t_{1-\kappa}\right)$ for a two-tail test from the null hypothesis of no impact. Because t-values are multiples of the standard error of an impact estimator, the minimum detectable effect is $\left(t_{\frac{\alpha}{2}}+t_{1-\kappa}\right)$ times the standard error. These critical t-values depend on the number of degrees of freedom. To achieve a statistical power κ and statistical significance of α it must be that the minimum detectable effect,

$$\text{MDE}=\left(t_{\frac{\alpha}{2}}+t_{1-\kappa}\right)*SE\left(\widehat{\beta}\right)$$

Formulas to compute MDE are well known and widely used by researchers. We refer readers to Bloom (2005) for detailed exposition of formulas for computing MDE or MDES. It can be shown that the MDE for a clustered randomized design with J groups of size n for a given power (κ), significance level (α) for a two-sided test, and proportion of subjects allocated to treatment group (P), inter-cluster correlation of ρ, and standard deviation of outcome measure (σ) is given by

$$MDE=(t_{\alpha/2}+t_{1-\kappa})*\sqrt{\frac{\rho}{P(1-P)J}+\frac{1-\rho}{P(1-P)J*n}}*\sigma$$

We also conduct power calculations using the actual compliance rates. If c is the share of subjects initially assigned to the treatment group who receive the treatment and s is the share of subjects initially assigned to the comparison group who receive the treatment, MDE is given by[14]

$$MDE=(t_{\alpha/2}+t_{1-\kappa})*\sqrt{\frac{\rho}{P(1-P)J}+\frac{1-\rho}{P(1-P)J*n}}*\frac{\sigma}{c-s}$$

[13] For a one-tail test, $t_{\alpha/2}$ is substituted by t_α.

[14] See Bloom (2005) and Duflo, Glennerster, and Kremer (2007) for further discussion.

Table 5
Ex-post minimum detectable effect sizes (MDES).

Country	Variable description	Inter-cluster correlation	Treatment effect	Control mean	Full Compliance			Actual Compliance		
					MDE	MDE (% Control mean)	MDES (Standard deviation)	MDE	MDE (% Control mean)	MDES (Standard deviation)
Bosnia	Asset value	0.033	−4388	111,229	22,940	21%	0.18	118,861	107%	0.92
	Profit (respondent business)	0.009	672	2903	1530	53%	0.18	7927	273%	0.92
	Any self-employment income	0.033	0.06	0.67	0.08	12%	0.18	0	63%	0.92
	Self-employment income (profit)	0.004	74	6122	1648	27%	0.18	8541	140%	0.92
	Wage income	0.028	323	6897	1578	23%	0.18	8176	119%	0.92
	Total consumption per capita	0.022	−648	4167	867	21%	0.18	4494	108%	0.92
	Food expenditure	0.031	−4	117	16	13%	0.18	80	69%	0.92
Ethiopia	Started business in last 3 years	0.015	−0.02	0.07	0.02	21%	0.08	0.06	84%	0.31
	Net revenues (profits)	0.000	526	146	333	227%	0.05	1320	903%	0.20
	Self-employment (profits)	0.000	513	755	346	46%	0.05	1374	182%	0.20
	Wage income	0.031	49	294	116	39%	0.10	459	156%	0.39
India	Assets (stock)	0.009	598	2498	1132	45%	0.09	8709	349%	0.66
	Profit	0.001	354	745	866	116%	0.07	6665	895%	0.56
	Started a business (12 months)	0.003	0.01	0.05	0.02	35%	0.08	0.13	270%	0.58
	Self-employment (profit)	0.001	354	745	866	116%	0.07	6665	895%	0.56
	Wage income	0.221	−526	2988	1028	34%	0.26	7907	265%	2.04
	Total consumption per capita	0.038	−12	525	35	7%	0.11	268	51%	0.85
	Food expenditure per capita	0.025	−9	84	14	17%	0.13	108	128%	0.96
Mexico	Profits in the last two weeks	0.000	0.11	145.47	166.20	114%	0.04	1445	994%	0.39
	Started a business (12 months)	0.006	−0.01	0.10	0.02	15%	0.05	0.13	132%	0.45
	household business income	0.014	58	840	186	22%	0.06	1621	193%	0.54
	HH wage income	0.020	−30	4541	346	8%	0.07	3013	66%	0.58
	Value of assets	0.016	−1534	8319	1438	17%	0.06	12,504	150%	0.54
	Amount spent on food (weekly)	0.040	4	874	50	6%	0.08	439	50%	0.73
Mongolia	Assets (stock)	0.065	−29	2236	692	31%	0.32	2694	120%	1.23
	HH business profit	0	−5	−27	16	−59%	0.23	62	−231%	0.89
	Respondent business profit	0	−8	−12	13	−105%	0.23	49	−408%	0.89
	Respondent started business	0.001	0.01	0.07	0.06	90%	0.23	0.23	350%	0.90
	Self-employment (profit)	0	−4.79	−26.85	15.94	−59%	0.23	62.02	−231%	0.89
	Wage income	0.012	−253	414	407	98%	0.25	1583	382%	0.96
	Total consumption per capita	0.115	0.11	10.95	0.21	2%	0.37	0.82	7%	1.44
	Monthly HH food expenditure	0.142	0.14	10.34	0.24	2%	0.40	0.92	9%	1.54
Morocco	Assets	0.129	1448	15,984	4421	28%	0.17	26,470	166%	1.05
	Profit	0.041	2005	9056	5547	61%	0.12	33,213	367%	0.71
	Has a self-employment activity	0.186	−0.02	0.83	0.06	7%	0.20	0.37	44%	1.21
	Wage income	0.049	447	27,669	6378	23%	0.18	38,192	138%	1.06
	Self-employment profit	0.041	2005	9056	5547	61%	0.12	33,213	367%	0.71
	Total monthly HH consumption	0.096	−46	3057	303	10%	0.16	1812	59%	0.94
	Monthly HH food consumption	0.138	3	1784	140	8%	0.18	836	47%	1.08
Philippines	Profits	0.00113	6281.7	14,198	8091	57%	0.21	11,238	79%	0.29
	Total Revenues	0.00246	5542.0	56,974	18,557	33%	0.21	25,773	45%	0.29

(*continued on next page*)

Table 5 (*continued*)

Country	Variable description	Inter-cluster correlation	Treatment effect	Control mean	Full Compliance MDE	MDE (% Control mean)	MDES (Standard deviation)	Actual Compliance MDE	MDE (% Control mean)	MDES (Standard deviation)
	Expenses	0.00052	114.5	41,497	15,533	37%	0.21	21,574	52%	0.29
	Food consumption	0.02102	789.3	8302	1225	15%	0.21	1701	20%	0.29
Uganda	Profits	0.00116	25.6	458	150	33%	0.15	366	80%	0.37
	Expenses	0.00491	−42.5	773	221	29%	0.15	540	70%	0.37
	Total consumption	0.01739	−79.7	587	93	16%	0.15	226	39%	0.36
	Food consumption	0.000124	319.5	55,393	6882	12%	0.15	16,785	30%	0.37

Notes: Notes: Clusters used in the original studies are used to compute inter-cluster correlations. Clusters include 14 bank branches for Bosnia, 66 Kebele for Ethiopia, 104 areas or communities in India, 238 clusters in Mexico, 162 villages in Morocco, 40 villages in Mongolia, 8 bank branches in the Philippines, and 37 towns in Uganda. Treatment effects are intention to treat (ITT) estimates of microfinance treatment. Minimum detectable effect (MDE) are expressed in the same units as the outcome variable. Minimum detectable effect size (MDES) are in standard deviation units. MDE (% Control mean) is MDE expressed as % of the mean of the outcome for the control group.

The multiplier $\left(t_{\frac{\alpha}{2}} + t_{1-\kappa}\right)$ is approximately equal to $(1.96 + 0.84 = 2.8)$ for two-tail tests given 80 percent statistical power and 5% significance level.

MDEs are expressed in the same units as the unit of the outcome measure. Since we present MDEs for several outcomes measured in different contexts, for comparability, we present minimum detectable effect size (MDES), which is measured in standard deviation units.

$$MDES = \frac{MDE}{\sigma} = \left(t_{\alpha/2} + t_{1-\kappa}\right) * \sqrt{\frac{\rho}{P(1-P)J} + \frac{1-\rho}{P(1-P)J*n}} * \frac{1}{c-s}$$

We also report MDE as percent of the mean of the outcome measure for the control group. For this we simply take the MDE computed using the equation above and express it as percentage of the control mean of the relevant outcome measure. MDE and MDES are computed for situations with 100 percent compliance as well as for actual compliance rates reported in each of the studies.

As an example, we show computations of MDE and MDES for profit outcome for the Uganda study. We compute MDE or MDES assuming 80% statistical power and 5% significance level for a two-tail test. Since the Uganda study had net compliance of 41%, $n = 1390, and P = 0.54$. For profit variable we have, $\sigma = 996.7$. Since the study was randomized at the individual level, we have $J = 1 and \rho = 0$. Plugging these numbers into the above equation, we get 79.9% MDE as percent of control mean and MDES of about 0.368 standard deviations.

$$MDE = (1.96 + 0.84) * \sqrt{\frac{0}{0.54(1-0.54)1} + \frac{1-0}{0.54(1-0.54)*1*1390}} * \frac{996.7}{0.41} = 366.3$$

Since the mean of profit for the control group is 458.4.

$$MDE \text{ as percent of control mean} = \frac{366.3}{458.4} * 100\% = 79.9\%$$

$$MDES = \frac{MDE}{\sigma} = \frac{366.3}{996.7} = 0.368$$

Note that there is no consensus on what would constitute a high MDE as this is contingent on expectations of what could reasonably be accomplished by a program. The initial hype about microfinance led some to expect incomes of participants to double in a short period of time. More sober expectations of programs that have a good internal rate of return generally achieve increases in income of 15–20% per year.

3.2. Compliance rates

All eight studies studied in this exercise suffer from imperfect compliance. Net compliance rates for any MFI loans are low for all eight studies and range between 6.9 ppts for Mexico and 43.9 ppts for Bosnia. The randomized designs used in each of these studies only influences the probability that someone receives a treatment. Even though these studies seem to have taken into consideration the possibility of imperfect compliance while determining sample sizes, actual compliance was much lower than expected. For example, the partner MFI for Banerjee et al. (2015) experiment in India expected that 80% of eligible households would borrow. However, only 26.7 percent of the eligible households borrowed from the partner MFI and the net compliance rate ended up being only about 13% for partner MFI loans.

Controlling for baseline covariates helps improve precision of the estimates and thus increases the chances of detecting any effects. The minimum detectable effect sizes (MDES) computed in the pooled exercise do not consider the fact that six of the individual studies can control for a broader set of baseline covariates and four of the studies control for baseline outcome values. However, in majority of the cases, MDES are so large that controlling for a complete set of baseline covariates would not improve the precision of the estimates by enough for the effects to be detected.

Results of power calculations for the individual studies are presented in Table 5. We focus on discussing MDES for variables in outcome groups for self-employment activities (profits, started a business, and engagement in self-employment), income and consumption. MDES under perfect compliance are generally reasonable, ranging between 12 and 53 percent of control mean (0.18 SD) in Bosnia, 21 to 227 percent (0.05 to 0.1 SD) in Ethiopia, 7 to 116 percent (0.07 to 0.26 SD) in India, 8 to 114 percent (0.04 to 0.08 SD) in Mexico, 2 to 105 percent (0.23 to 0.4 SD) in Mongolia, 8 to 61 percent (0.12 to 0.2 SD) in Morocco, 15 to 57 percent (0.21 SD) in the Philippines, and 12 to 33 percent (0.15 SD) in Uganda. Very high MDES value for profits in Ethiopia is due to high root mean squared error. However, extremely low net take-up rates (or low compliance rates) for microfinance products resulted in extremely large MDES. MDES for some variables are as high as 994% of the control group mean or as high as 1.23 standard deviations. As expected MDES are largest for profits from self-employment activities for most of the studies.

Under actual compliance rates, MDES for profits are as large as 273% of mean outcome for control group (0.92 SD) in Bosnia, 903% (0.2 SD) in Ethiopia, 895% (0.56 SD) in India, 994% (0.54 SD) in Mexico, 408% (0.89 SD) in Mongolia, 367% (0.71 SD) in Morocco, 79% (0.29 SD) in the Philippines, and 80% (0.37 SD) in Uganda. For consumption, MDES are 108% of mean outcome for control group (0.92 SD) in Bosnia, 51% (0.85 SD) in India, 50% (0.73 SD) in Mexico, 7% (1.44 SD) in Mongolia, 47% (1.08 SD) in Morocco, 20% (0.29 SD) in the Philippines, and 39% (0.36 SD) in Uganda. These large MDES, mostly a result of extremely low compliance rates, could explain the null results for many outcomes in these microfinance studies.

We also computed minimum detectable effect sizes for the combined sample in Table 6 (weighted) and appendix Table A4 (unweighted). For the combined sample, we assume clustering at the bank branch level for the Bosnia and the Philippines studies, both of which do not follow a clustered design but instead randomize at the individual level. The Uganda study is also randomized at the individual level and we assume clustering at the district level.

To compute cluster or group size we took the total sample size in the combined data and divided it by the total number of clusters in all eight studies. MDES for combined sample are still very high as a percentage of control means. This probably explains why we get insignificant treatment effects for some outcome measures even with the combined sample. For the combined sample, MDES are 10% of control mean (0.12 SD) for engagement in self-employment, 14% (0.04 SD) for starting a business in the near past, 44% (0.05 SD) for profits, and 13% (0.11 SD) for total consumption when we assume full compliance. With the 25% net compliance, more-or-less the average net compliance rates for the eight studies, the MDES increase by a factor of four.

Banerjee et al. (2015) make it clear that generating sufficient statistical power is a challenge for randomized evaluations of microcredit, and one of the main caveats of the all eight studies considered in this meta-analysis is poor statistical power resulting from low net take-up of microcredit. Many of the null results in these studies are within the confidence intervals that contain economically meaningful effect sizes of increased access to microcredit. This motivates them to suggest both additional studies

Table 6
Minimum detectable effect sizes for the pooled sample (Weighted).

Variable description	Intra-cluster correlation	Treatment effect	Standard Deviation	Control mean	Full Compliance: Minimum detectable effect (MDE)	Full Compliance: MDE (% Control mean)	Full Compliance: MDES (Standard deviation)	25% Compliance: Minimum detectable effect (MDE)	25% Compliance: MDE (% Control mean)	25% Compliance: MDES (Standard deviation)
Program MFI loan	0.21	0.32	0.43	0.06	0.04	63%	0.11	0.16	252%	0.42
Any loan	0.22	0.12	0.49	0.51	0.05	9%	0.11	0.19	37%	0.44
Program MFI loan amount	0.14	240.79	755	121.72	62.40	51%	0.09	249.62	205%	0.36
Any loan amount	0.07	397.02	6961	1412.39	456.01	32%	0.07	1824.05	129%	0.27
Assets	0.03	0.04	0.96	−0.08	0.05	−64%	0.05	0.19	−258%	0.21
Wage income (2 weekly)	0.07	−124.98	2662	387.05	170.77	44%	0.07	683.09	176%	0.26
profits	0.04	18.68	494	56.35	24.63	44%	0.05	98.50	175%	0.21
Total income	0.10	−96.31	2522	433.59	186.72	43%	0.08	746.87	172%	0.30
Revenues	0.11	25.28	1282	246.87	93.51	38%	0.08	374.05	152%	0.32
Expenses	0.08	8.31	986	146.73	62.74	43%	0.07	250.95	171%	0.27
Investment	0.00	5.56	215	8.43	9.65	115%	0.04	38.62	458%	0.18
Self-employment activity	0.28	0.03	0.50	0.50	0.05	10%	0.12	0.20	40%	0.48
Started a business	0.01	0.01	0.28	0.08	0.01	14%	0.04	0.05	56%	0.17
Total consumption	0.22	20.42	367	275.83	36.56	13%	0.11	146.22	53%	0.45
Food Consumption	0.18	3.03	144	150.53	12.17	8%	0.10	48.66	32%	0.39
Durable Consumption	0.05	5.58	145	45.22	9.57	21%	0.07	38.28	85%	0.27
Temptation goods	0.07	−1.35	33	22.00	2.12	10%	0.07	8.46	38%	0.27

Notes: Clusters used in the original studies are used to compute inter-cluster correlations. Clusters include 14 bank branches for Bosnia, 66 Kebele for Ethiopia, 104 areas or communities in India, 238 clusters in Mexico, 162 villages in Morocco, 40 villages in Mongolia, 8 bank branches in the Philippines, and 37 towns in Uganda. Treatment effects are intention to treat (ITT) estimates of microfinance treatment. Minimum detectable effect (MDE) are expressed in the same units as the outcome variable. Minimum detectable effect size (MDES) are in standard deviation units. MDE (% Control mean) is MDE expressed as % of the mean of the outcome for the control group.

and formal meta-analyses to better understand the impact of microfinance.

Five out of eight individual papers considered in this analysis are forthright about one common major caveat in their papers: imprecisely estimated null effects even when the effect sizes are economically meaningful. The Ethiopia study speculates that the reason for the failure to identify statistically significant impacts on key outcomes such as net revenues or livestock ownership is likely insufficient statistical power or measurement error.

The India paper notes that only a small difference in microfinance take-up between treatment and control areas means that the power and precision of the estimates are significantly lowered. Mexico paper notes that many of the statistically insignificant intent-to-treat estimates are economically meaningful effect sizes. They attribute the lack of precise nulls, even with a relatively large sample size, to a combination of the modest take up differential between treatment and control areas, heterogeneous treatment effects, and high variance and measurement error in outcomes.

To increase take-up rates among the treatment group and thus increase statistical power, the Mongolia paper offered credit to women who had expressed an initial interest in borrowing. Despite this attempt to increase statistical power, the paper still documents some quite substantial but imprecisely estimated impacts. The Morocco paper also designed and implemented a sampling strategy that would improve power to estimate the impact on borrowers, as well as to capture impacts representative at the village level. Still, they encounter issue of statistical power due to extremely low microfinance take-up rate differential in treatment and control areas. They do find that the impact on those who are more likely to borrow and those who actually borrow are much larger.

4. What do these studies and other studies of microfinance really tell us?

The eight papers we look closely at here use random assignment of microfinance to answer the question: What is the impact of increased access to microfinance? We believe the evidence presented in these papers is not conclusive enough to provide a definitive answer to suggest null impact of increased access to microfinance. While most find large estimated coefficients, few are statistically significant. The null results in these studies are mostly imprecisely estimated insignificant results for which the researchers cannot rule out the treatment effects with confidence. Pooling the data improves power, but still does not solve the issue. Bayesian hierarchical modeling offers a way to look closer at the data but does not solve problems with the designs of these studies.

The authors attribute highly imprecisely estimated null effects to a combination of the modest take-up differential between treatment and control areas, heterogeneous treatment effects, and high variance and measurement error in outcomes. These studies are, thus, not able to reject the null of no impact, even for some substantially large point estimates that are economically meaningful.

Even in rural Morocco, where the baseline access to credit is relatively low and there are no other sources of formal credit in study areas, take-up of credit is very low. The low net take-up of credit also made it difficult for the authors to reject economically significant but statistically insignificant null results. The problem of low take-up differential between treatment and control areas was most severe in urban India, where households had access to other sources of credit. In India, access to MFI loan increases by a modest 8.4 percentage points in treatment areas relative to control areas, but there was no statistically significant increase in access to any loan type in treatment areas.

Outcomes such as profits and revenues are hard to measure and are also characterized by large variances. Thus, measurement errors in these outcome variables also makes it difficult to get precise estimates of the impact of microfinance on these outcomes.

Some evidence from the eight randomized studies of microfinance also suggests that the impact of microfinance is not likely to be homogenous across the population or the prospective borrowers, which also makes it harder to detect impact on the average borrower. In the Ethiopia study, Tarozzi et al. (2015) find that areas that were assigned microfinance treatment saw overall increases in earnings from self-employment activities, mostly affecting the right tail of the distributions. Banerjee et al. (2015a) find in India that the impacts on income-generating activities are concentrated in the upper tail of the distribution. They also find that those who choose to borrow are more likely to expand their existing businesses or to start female-owned businesses. Small business investment and profits of preexisting businesses increased, but new businesses were not that profitable.

In the Mexico study, Angelucci et al. (2015), examines the extent of heterogeneous effects of microfinance by estimating quantile treatment effects and show stronger effects at the upper end of the distribution for revenues and profits. However, they do not find any noticeable pattern across the distribution for most outcomes.

In Morocco, Crépon et al. (2015) find significant increase in total self-employment profit, although there also appears to be great heterogeneity in these effects[15]. As in other studies, they find that the effect on profits is significantly positive at the higher quantiles but significantly negative at the lower quantiles of profitability. They also find that there are proportionally large average impacts on self-employment investments, sales, and profits among those who take-up microfinance.

It is also important to note that most of the studies (with exception of Morocco and Ethiopia) were conducted in settings where access to credit were already high, which means that these studies are likely capturing marginal borrowers. Wydick (2016) shows that the impacts of microfinance on marginal borrowers are likely to be lower than those on inframarginal borrowers. Thus, impact estimates obtained from experimental designs focusing on marginal microfinance borrowers in areas already served by microfinance are likely to understate the average treatment effects of microfinance. Ogden (2016) presents a non-technical discussion of eight microfinance experiments, focusing on the fact that existing studies are measuring microcredit expansion to additional or marginal clients, and so modest results should be expected. He also discusses the need for creative approaches to microcredit designs, for which there is little quantitative evidence.

Other, more recent studies have produced evidence for a positive impact of non-traditional microfinance. They suggest that a more creative approach to finance can lead to positive impacts for people. The success of microfinance is likely very context dependent and may depend on alleviating additional constraints that operate alongside credit constraints. For instance, Field (2013) find that relaxing constraints to credit structure can have positive impacts. The study uses a field experiment to compare the classic microfinance contract, which requires that repayment begin immediately after loan disbursement, to a microfinance contract that includes a two-month grace period as well as less frequent loan-repayment schemes. They find that the provision of a grace period and less frequent repayment schemes increases short-run business investment and long-run profits.

Burke, Bergquist and Miguel (2017) also look at a modified model of finance and find that the timing of when the loan is given can matter. Finance given when farmers have harvested, rather

[15] It should be noted that a recent paper by Bédécarrats et al. (2019) finds significant problems with this study and are unable to replicate the results.

than before planting as is the common approach, offers farmers inter-temporal arbitrage opportunities. This arbitrage leads to significant increases in farm profits and returns on investment.

It is possible that who receives the microfinance may matter for impacts. Most studies on microfinance, including the eight studies we discuss here, look predominately at finance provided to women. In an experimental study of microloans, cash grants and business skills training where 40% of the sample are men, Fiala (2018) finds that male-owned microenterprises that are provided both access to loans and training report significantly higher profits in the short-run. These loans were also subsidized so that interest rates were 20% rather than the normal 27%. Thinking critically about the model and participants of microfinance programming could prove impactful for people.

5. Next steps for microfinance and research

The results from our analysis of the eight studies discussed here suggests that the current evidence for the impacts of microfinance is very weak and has likely been misinterpreted by critics of microfinance. Clearly, researchers have not been able to find massively transformative impacts from microfinance. While none of the studies discussed here have the statistical power necessary to identify modest impacts, we are able to detect impacts on business profits and revenues when the data from the studies are combined together. But this result must be taken cautiously due to low power even in the pooled sample.

What can be done to improve the impact of microfinance? Field, Hollander and Pande (2014) conduct a review of microfinance literature and suggest five ways to better the microfinance model. These suggestions include building more flexibility into the microfinance contract (as in Field, 2013; Barboniy and Agarwal, 2018 and Battagliay, Gulesci and Madestam, 2018), directly encouraging greater business investment (perhaps through training like in Fiala, 2018), using microfinance to build social capital, anticipating and measuring a broader range of development outcomes, and focusing more on the rural population. We believe these are all excellent suggestions that have paid off in the recent literature.

Improving the evidence for microfinance is ultimately based on the need for better designed studies. We end by discussing what an ideal experimental design might look like. This discussion is based on many of the issues we raised earlier, as well as additional ideas from the literature. While experimentally evaluating microfinance programs that have low take-up rates and many competitors is challenging, we believe it is not impossible.

One of the major limitations with many of the studies we discuss is that most utilize an encouragement design. Power challenges are not necessarily inherent to encouragement designs. However, encouragement is unpredictable. Future studies that randomize at the cluster level will need to do a better job of identifying who is likely to take-up loans *ex ante*. This will both improve power and ensure the ITT estimates are closer to actual impact sizes.

Researchers should also consider collecting outcome data at multiple periods. As McKenzie (2012) suggests, taking multiple measurements of noisy outcomes such as business profits, income and expenditures at relatively short intervals allows for averaging out noise, increasing power significantly.

It is also necessary to identify contexts where households are truly credit constrained and borrowing is not frowned upon by local customs. One of the reasons for low take-up of loans even in rural Morocco, where the availability of other credit sources is extremely limited, could be because traditional borrowing models are frowned upon by most followers of Islam. The product must be truly appreciated, or else there will be nothing meaningful to measure.

Of course, the best power situation is obtained when randomization is at the individual level. A few of the studies we discuss above were able to do this. This greatly improves power, but at the cost of increasing the likelihood of spillovers. It is also not feasible in most circumstances, either due to resource constraints, concerns about control individuals simply finding other finance options, or microfinance institutions not willing to turn down eligible applicants.

Researchers also need to be clear about what exactly is being evaluated. Many of the studies here can be described as measuring the expansion of microfinance to either new areas or marginal clients. When individuals have other options for microfinance, it is possible that researchers are simply comparing one microfinance option against other credit options.

It is also important to realize that even if liquidity constraints are a binding constraint for income growth, there may be other constraints that need to be loosened for microfinance to deliver on its promise. For example, providing business training may be important for encouraging poor households to take-up self-employment activities and make them profitable. Fiala (2018) finds male-owned microenterprises that are provided both access to loans and training report significantly higher profits. Thus, microfinance interventions combined with other interventions, like business or skills training, may improve the chances of finding impacts.

The impacts of microfinance are also likely heterogeneous. More studies are needed that allow for estimating meaningful heterogenous effects. This could mean conducting a rich baseline with a relatively large sample size or doing high quality qualitative and exploratory work before conducting an experiment on a product.

What do these results mean for policy makers? The most important take-away we have is that the existing experimental evidence is not of sufficient quality to determine whether microfinance has positive impacts on people's lives. It appears unlikely that it has transformative impacts, but we can't be sure it does not have some positive impact. Note that the size of the coefficients we estimate in the pooled sample are quite large and dwarf many other development intervention effects. While we must be careful in interpreting this finding, it suggests that there is value to studying the question closely. It also suggests that there could be value for donors and governments to subsidize microfinance, as was originally suggested by Mohammed Yunus.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Chris Blattman, David McKenzie, Rachael Meager and Jörg Peters for their helpful comments. Funding for research assistant time for this study was provided by Opportunity International. The results and opinions are those of the authors only.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.worlddev.2019.104773.

References

Angelucci, Manuela, Karlan, Dean, & Zinman, Jonathan (2015). Microcredit impacts: Evidence from a randomized microcredit program placement experiment by

Compartamos Banco. *American Economic Journal: Applied Economics, 7*(1), 151–182.

Attanasio, Orazio, Augsburg, Britta, De Haas, Ralph, Fitzsimons, Emla, & Harmgart, Heike (2015). The impacts of microfinance: Evidence from joint-liability lending in Mongolia. *American Economic Journal: Applied Economics, 7*(1), 90–122.

Augsburg, Britta, De Haas, Ralph, Harmgart, Heike, & Meghir, Costas (2015). The impacts of microcredit: Evidence from Bosnia and Herzegovina. *American Economic Journal: Applied Economics, 7*(1), 183–203.

Banerjee, Abhijit, Duflo, Esther, Glennerster, Rachel, & Kinnan, Cynthia (2015a). The miracle of microfinance? Evidence from a randomized evaluation. *American Economic Journal: Applied Economics, 7*(1), 22–53.

Banerjee, Abhijit, Duflo, Esther, Goldberg, Nathanael, Karlan, Dean, Osei, Robert, Pariente, William, Shapiro, Jeremy, Thuysbaert, Bram, & Udry, Christopher (2015b). A multifaceted program causes lasting progress for the very poor: Evidence from six countries. *Science, 348*, 6236.

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

Bangdiwala, Shrikant I. et al. (2016). Statistical methodologies to pool across multiple intervention studies. *Translational behavioral medicine, 6*(2), 228–235.

Bédécarrats, Florent, Guérin, Isabelle, Morvant-Roux, Solène, & Roubaud, François (2015). Estimating microcredit impact with low take-up, contamination and inconsistent data. A replication study of Crépon, Devoto, Duflo, and Pariente (American Economic Journal: Applied Economics, 2015). *International Journal for Re-Views in Empirical Economics, 3*.

Bloom, Howard S. (2005). Randomizing groups to evaluate place-based programs. *Learning more from social experiments: Evolving analytic approaches*, 115–172.

Burke, Marshall, Betgquist, Lauren Falcao, & Miguel, Edward (2017). Selling low and buying high: An arbitrage puzzle in Kenyan villages. *Working paper*.

Marianna, Battaglia, Selim, Gulesci, & Andreas, Madestam (2018). Repayment Flexibility and Risk Taking: Experimental Evidence from Credit Contracts. *CEPR Discussion Papers 13329, C.E.P.R. Discussion Papers*.

Giorgia, Barboni, & Agarwal, Parul (2018). *Knowing what's good for you: Can a repayment flexibility option in microfinance contracts improve repayment rates and business outcomes?. Working paper*. Princeton University.

Crépon, Bruno, Devoto, Florencia, Duflo, Esther, & Parienté, William (2015). Estimating the impact of microcredit on those who take it up: Evidence from a randomized experiment in Morocco. *American Economic Journal: Applied Economics, 7*(1), 123–150.

Duflo, Esther, Glennerster, Rachel, & Kremer, Michael (2007). Using randomization in development economics research: A toolkit. *Handbook of development economics, 4*, 3895–3962.

Fiala, Nathan (2018). Stimulating microenterprise growth: Results from a loans, grants and training experiment in Uganda. *Working paper*.

Field, Erica et al. (2013). Does the classic microfinance model discourage entrepreneurship among the poor? Experimental evidence from India. *American Economic Review, 103*(6), 2196–2226.

Field, Erica, Abraham J. Hollander, and Rohini Pande. "Micro finance: Points of Promise." (2014).

Gelman, Andrew, & Carlin, John (2014). Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors. *Perspectives on Psychological Science, 9*(6), 641–651.

Ioannidis, John, & Stanley, Tom D. (2017). and Hristos Doucouliagos. "The power of bias in economics research.". *The Economic Journal, 127*, 605.

Karlan, Dean, & Zinman, Jonathan (2011). Microcredit in Theory and Practice: Using Randomized Credit Scoring for Impact Evaluation. *Science, 332*(6035), 1278–1284.

McKenzie, David (2012). Beyond baseline and follow-up: The case for more T in experiments. *Journal of development Economics, 99*(2), 210–221.

McKenzie, David, & Woodruff, Christopher (2013). What are we learning from business training and entrepreneurship evaluations around the developing world? *The World Bank Research Observer, 29*(1), 48–82.

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

Morton SC, Adams JL, Suttorp MJ, Shekelle PG (2004) Meta-regression approaches: what, why, when, and how?, Technical Review 8, Agency for Healthcare Research and Quality Publication No. 04–0033. Rockville.

Ogden, Timothy. "The Case for Social Investment in Microcredit." Financial Access Initiative. 2016

Pitt, M. M., & Khandker, S. R. (1998). The impact of group-based credit programs on poor households in Bangladesh: Does the gender of participants matter? *Journal of political economy, 106*(5), 958–996.

Romano, Joseph P., & Wolf, Michael (2005). Stepwise multiple testing as formalized data snooping. *Econometrica, 73*(4), 1237–1282.

Tarozzi, Alessandro, Desai, Jaikishan, & Johnson, Kristin (2015). The impacts of microcredit: Evidence from Ethiopia. *American Economic Journal: Applied Economics, 7*(1), 54–89.

Wydick, Bruce (2016). Microfinance on the margin: Why recent impact studies may understate average treatment effects. *Journal of Development Effectiveness, 8*(2), 257–265.