Science
AAAS

Microcredit in Theory and Practice: Using Randomized Credit Scoring for Impact Evaluation

Dean Karlan, *et al.*

Science **332**, 1278 (2011);
DOI: 10.1126/science.1200138

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.

Microcredit in Theory and Practice: Using Randomized Credit Scoring for Impact Evaluation

Dean Karlan[1,2,3]* and Jonathan Zinman[2,3,4]*

Microcredit institutions spend billions of dollars fighting poverty by making small loans primarily to female entrepreneurs. Proponents argue that microcredit mitigates market failures, spurs micro-enterprise growth, and boosts borrowers' well-being. We tested these hypotheses with the use of an innovative, replicable experimental design that randomly assigned individual liability microloans (of $225 on average) to 1601 individuals in the Philippines through credit scoring. After 11 to 22 months, we found evidence consistent with unmet demand at the current price (a roughly 60% annualized interest rate): Net borrowing increased in the treatment group relative to controls. However, the number of business activities and employees in the treatment group decreased relative to controls, and subjective well-being declined slightly. We also found little evidence that treatment effects were more pronounced for women. However, we did find that microloans increase ability to cope with risk, strengthen community ties, and increase access to informal credit. Thus, microcredit here may work, but through channels different from those often hypothesized by its proponents.

Microcredit—broadly speaking, the provision of small loans (typically $100 to $500 U.S.) to very small businesses, typically self-run enterprises with few if any employees—is an increasingly common weapon in the fight to reduce poverty and promote economic growth (*1*). The motivation for the continued expansion of microcredit, or at least for the continued flow of subsidies to both nonprofit and for-profit lenders, is the presumption that expanding credit access is a relatively efficient way to fight poverty and promote growth. Yet despite strong claims about the effects of microcredit on borrowers and their businesses (e.g., the 2006 Nobel Peace Prize to Muhammad Yunus and the Grameen Bank), there is relatively little rigorous evidence about these programs (*2*).

In practice, the policy discussion about microlending typically emphasizes the upside, arguing that microlending mitigates market failures (by making access to credit possible for entrepreneurs previously unable to get credit), empowers women (e.g., improves their decision-making power by giving them more financial independence), and spurs enterprise growth and improves subjective well-being (e.g., increase in life satisfaction, self-esteem, and optimism; decrease in level of stress).

In theory, expanding credit access may not have positive effects on borrowers and could even have negative effects. Financial institutions may disrupt relatively efficient "informal" (community- or family-based) mechanisms (*3*). The often high cost of microcredit, from 10% annualized interest rates to 100%, means that even higher returns to capital are required for microcredit to produce improvements in business income, and thus in household income and consumption. Finally, some argue that psychological biases may induce some to "overborrow" and do themselves more harm than good (*4*).

"Traditional" microlenders target women who operate small-scale businesses and use group lending mechanisms (*5*). But as microlending has expanded and evolved into its "second generation," it often ends up looking more like traditional retail or small-business lending: For-profit lenders extend individual liability credit in increasingly urban and competitive settings. For example, recent estimates suggest that about one-half of microfinance institutions are individual liability lenders, and about one-quarter are for-profits or cooperatives (*6*–*9*).

We have conducted a randomized evaluation of second-generation microcredit by working with First Macro Bank (FMB) to implement a novel, replicable experimental design that uses credit scoring to randomly assign individual liability loans as a source of exogenous variation in credit access (*10*–*12*). FMB is a for-profit lender that makes small, 3-month loans at 60% annualized interest rates to micro-entrepreneurs in the outskirts of Manila and receives technical assistance from a U.S. Agency for International Development (USAID) contractor (*13*). Nonrandomized empirical evaluations of microcredit impacts are typically complicated by classic endogeneity problems: Client self-selection and lender strategy likely produce correlations among credit access, ultimate outcomes, and critical unobserved inputs (e.g., client opportunity sets, preferences, and aptitude) that confound attempts to make causal inferences of microcredit impacts (*14*).

We worked with the lender to build a quantitative model that distinguishes creditworthy or uncreditworthy applicants from marginal ones. Marginal applicants then get approved for a loan according to some preassigned probability. This method provides lenders with a way to take systematic, controlled risks when refining underwriting strategies. It also provides researchers and policymakers with a source of exogenous variation in access to credit that may be used, in conjunction with follow-up data (e.g., on business and household outcomes), to help identify the impacts of microcredit from a change in the screening criteria of existing lenders on marginal applicants. Note that impacts may differ for inframarginal applicants; we discuss this and other external validity issues below. Nonetheless, our methodology is transferable to many different types of lenders and settings.

The ability to transfer an evaluation method to a range of contexts is particularly important given the unsettled state of evidence on microfinance impacts. Prior studies have used various methodologies to address endogeneity problems and have found varied impacts or lack thereof (*15*–*23*). Is the variation in estimated impacts across studies due to methodology, to true underlying heterogeneity in borrower characteristics and market conditions, or to both? In particular, we draw the reader's attention to (*2*), which provides a summary of results and methodological issues of several nonexperimental impact evaluations to date. Applying similar experimental methodological approaches across different settings will help us to address these issues and paint a more complete picture for theory testing and policy evaluation.

Setting. Our cooperating lender, FMB, has operated as a rural bank in the metro Manila region of the Philippines since 1960. Filipino "microlenders" include both for-profit and nonprofit lenders offering small loans to micro-entrepreneurs (average $220) that are short-term (less than 1 year), uncollateralized, and on fixed schedules of equal periodic repayments.

Most Filipino microlenders operate on a small scale relative to microfinance institutions (MFIs) in the rest of Asia (*24*), and our lender is no exception. FMB maintained a portfolio of about 1400 individual and 2000 group borrowers throughout the course of the study. This portfolio represents a small fraction of its overall lending, which also includes larger business and consumer loans as well as home mortgages.

Microloan borrowers typically lack the credit history and/or collateralizable wealth needed to borrow from traditional institutional sources such as commercial banks. This holds for our sample—which is only marginally creditworthy by FMB's

[1]Department of Economics, Yale University, New Haven, CT 06520, USA. [2]Innovations for Poverty Action, 101 Whitney Avenue, New Haven, CT 06510, USA. [3]Jameel Poverty Action Lab, Massachusetts Institute of Technology, Cambridge, MA 02142, USA. [4]Department of Economics, Dartmouth College, Hanover, NH 03755, USA.

*To whom correspondence should be addressed. E-mail: dean.karlan@yale.edu (D.K.); jzinman@dartmouth.edu (J.Z.)

standards—even though our subjects have average income and education levels. Table 1 provides some demographics on our sample frame relative to the rest of Manila and the Philippines.

Casual observation suggests that many micro-entrepreneurs in our study population face binding credit constraints. Credit bureau coverage of micro-entrepreneurs in the Philippines is quite thin, so building a credit history is difficult for many business owners and consumers. Informal credit markets are robust, and serial borrowing from moneylenders charging 20% or more per month is common (more than 30% of our sample reported borrowing from moneylenders during the past year). Trade credit (i.e., credit from suppliers) is quite uncommon.

The loan terms granted in this experiment were the lender's standard ones for first-time borrowers. Loan sizes ranged from 5000 to 25,000 pesos, which is substantial relative to borrower income. For example, the median loan size made under this experiment, 10,000 pesos ($220), was 37% of the median borrower's net monthly income. Loan maturity was 13 weeks, with weekly repayments. The monthly interest rate was 2.5%, charged over the declining balance. Several upfront fees combined with the interest rate to produce an effective annual interest rate greater than 60% (*25*).

The lender conducted underwriting and transactions in its branch network. At the onset of this study, FMB changed its risk assessment process from one based on weekly credit committee meetings to one that used computerized credit scoring.

Delinquency and default rates were substantial. One-third of the loans in our sample were paid late at some point, and 7.4% were charged off.

Methods. We drew our sample frame from the universe of several thousand applicants who applied at eight of the lender's nine branches between 10 February 2006 and 16 November 2007 (*26*). The branches were located in the provinces of Rizal, Cavite, and the National Capital Region. The lender maintained normal marketing procedures by having account officers canvass public markets and hold group information sessions for prospective clients, and these sessions did not include explicit discussion of the credit-scoring procedures.

Our research design uses credit-scoring software to randomize the approval decision for marginally creditworthy applicants (although "marginal" in this case actually encompassed 74% of the sample frame), and then uses data from household/business follow-up surveys to measure impacts on credit access and on several classes of more ultimate outcomes of interest. The survey data are collected by a firm, hired by the researchers, that has no ties to the lender.

Table 1 provides some summary statistics, from ex ante application data, on our sample frame of 1601 marginally creditworthy applicants, nearly all (1583) of whom were first-time applicants to the lender. Our sample was largely female, with a typical household size and with educational attainment and household income in line with averages for Metro Manila. The most common business was a sari-sari (small grocery/convenience) store. Other common businesses were food vending and various services (e.g., auto and tire repair, water supply, tailoring, and barbers and salons).

The lender identified marginally creditworthy applicants by means of a proprietary credit-scoring algorithm developed collaboratively between the lender and the researchers, based on business capacity, personal financial resources, outside financial resources, personal and business stability, and demographic characteristics (*27*). Credit bureau coverage of our study population is very thin, and our lender does not use credit bureau information as an input into its scoring. Scores ranged from 0 to 100, with applicants scoring below 31 rejected automatically and applicants scoring above 59 approved automatically. Our 1601 marginally creditworthy applicants fell into two randomization "windows": low (scores 31 to 45, with

Table 1. Demographics. Sample frame data taken from lender's application data unless otherwise noted. Per capita figures for Manila and the Philippines assume an average household size of 5.0 people. Source: www.census.gov.ph/data/quickstat/index.html.

	Our sample frame						Metro Manila		Philippines	
	All		Applicants with 60% chance of approval		Applicants with 85% chance of approval					
	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Applicant is female	85%	—	86%	—	85%	—				
Applicant is married	78%	—	53%	—	82%	—				
Age of applicant	42.1	42.0	41.8	42.0	42.1	42.0				
Education level of applicant*										
Elementary	11%	—	19%	—	10%	—	12%		33%	
High school	44%	—	49%	—	43%	—	42%		37%	
Postsecondary or college	45%	—	32%	—	47%	—	47%		31%	
Household size	5.1	5.0	5.0	5.0	5.1	5.0	5.0		5.0	
Number of dependents	2.28	2	2.29	2	2.28	2				
Applicant owns a sari-sari (corner) store	49%	—	55%	—	48%	—				
Monthly household income (Filipino pesos)†	24,920	17,245	19,524	14,150	25,826	17,800	25,917	18,333	14,417	9250
Monthly household income per capita (Filipino pesos)	5301	3540	4193	3191	5488	3569	5183	3667	2883	1850
Number of businesses owned by household	1.15	1	1.20	1	1.14	1				
Applicant's business has employees	25%	—	17%	—	26%	—				
Number of loans with First Macro Bank during study	2.0	2	2.3	2	2.0	2				

*Education data on sample frame are from the follow-up survey, where 97% of the sample frame is aged 20 to 59. Education data on Manila and the Philippines, restricted to Filipinos aged 20 to 59, are from www.census.gov.ph/data/sectordata/2003/fl03tabA.htm. †Monthly household income data on sample frame taken from the following questions from the follow-up survey: "How much was the total income (including remittances) earned by your household in the past month (gross calculation before expenses)?" less the sum total of "How much did each household business spend on each of the following categories of business expenses during the past month: [inventory, utility bills, wages and salaries for helpers, rent for machinery and equipment, rent for building and land, taxes, maintenance and general repairs, business-related transportation, and other expenses]?" Monthly household income data on Manila and the Philippines are from www.census.gov.ph/data/sectordata/2006/fies0607r.htm, where, according to the Family Income and Expenditures Survey, "total family income includes primary income and receipts from other sources received by all family members … and net receipts derived from the operation of family-operated enterprises/activities."

60% probability of approval) and high (scores 46 to 59, with 85% probability of approval). The randomization was opaque to loan officers, their direct supervisors, and applicants, in the sense that none of these parties saw actual credit scores, knew the details of the algorithm, or knew there was a random component to application decisions. In total, 1272 applicants were assigned to the treatment (loan approval) group, leaving 329 in the control (loan rejection) group.

The motivation for experimenting with credit access on a pool of marginal applicants is twofold. First, it focuses on those who are targeted by initiatives to expand access to credit, because those with higher credit scores are likely to have easier access to credit in general. Second, (randomly) approving some marginally creditworthy applicants generates data points on the lender's profitability frontier (by taking controlled risks). This feeds into revisions to the credit-scoring model and helps improve lender profits.

Table S1 provides some confirmation of two key conditions needed for our study design to produce exogenous variation in access to credit. First, the randomization was implemented properly: Pretreatment characteristics did not predict assignment to treatment in the full sample, in the sample that completed the follow-up survey, for females, or for men. Second, treatment assignment did not influence follow-up survey completion, which was not predicted by assignment to treatment, nor by interaction of assignment to treatment and baseline covariates (*28*).

Our sample frame and treatment assignments were created in the flow of the lender's three-step credit-scoring process. This process is replicable because it is relatively easy to administer operationally. Moreover, it can be augmented to introduce random assignment into other elements of loan contracting besides the approve/reject decision: pricing, loan amount, maturity, etc.

First, loan officers screened potential applicants on FMB's "basic four requirements" (18 to 60 years old, in business for at least 1 year, in residence for at least 1 year if homeowner or at least 3 years if renter, and daily income of at least 750 pesos); 2158 applicants passed this screen. Second, loan officers entered household and business information on these 2158 applicants into the credit-scoring software, and the software then rendered its application disposition within seconds. Of this group, 391 applications received scores in the automatic approval range, And 166 applications received scores in the automatic rejection range. The remaining 1601 applicants had scores in one of the two randomization windows (approve with 60% or 85% probability), and this group constituted our sample frame. Of these 1601 marginal applicants, 1272 were assigned "approve" and 329 applicants were assigned "reject" by the software, which simply instructed loan officers to approve or reject; that is, it did not display the application score or make any mention of the randomization. Neither loan officers, branch managers, nor applicants were informed about the credit-scoring algorithm or its random component.

The credit-scoring software's decision was contingent on complete verification of the application information, so the third step involved any additional due diligence deemed necessary by the loan officer or his supervisor. Verification steps included visits to the applicant's home and/or business, meeting with neighborhood officials, and checking references (e.g., from other lenders). If loan officers found discrepancies, they updated the information in the credit-scoring software, and in some cases the software changed its decision from approve to reject. In other cases, applicants decided not to go forward with completing the application, or completed the application successfully but did not avail the loan.

In all, there were 351 applications assigned out of the 1272 assigned to treatment that did not ultimately result in a loan. Conversely, there were five applications assigned to the control (rejected) group that did receive a loan (presumably because of loan officer noncompliance or clerical errors).

In all cases, we used the original treatment assignment from step 2 to estimate treatment effects; that is, we used the random assignment (loan approval or rejection) to estimate intention-to-treat effects: the treatment effect on all randomly assigned to receive credit, irrespective of whether the bank complied with the random assignment. One alternative, the treatment on the treated (TOT), would instead scale the intention-to-treat estimate up by the reciprocal of the compliance rate, and then present the estimated treatment effect on those who do ultimately receive the treatment.

Table 2. Microcredit in theory: Intention-to-treat effects of credit access on widely hypothesized outcomes. For the full sample and for females, data are OLS results for the independent variable "assigned a loan"; Huber-White SEs and control group means for the dependent variable listed in each row are also shown. The incremental effect on males is shown as an estimate for the interaction between "assigned a loan" and "male." Variation in sample sizes is due to survey question nonresponse. The summary index is in standard deviation units of the average outcome of its components. All estimates control for probability of assignment to treatment and for timing of treatment assignment and survey measurement. Borrowing measures do not count the 1% of loans that are too large (>50,000 pesos) to be plausibly affected by the treatment.

	Full sample			Females			Incremental effect on males		
	OLS result	SE	Control group mean	OLS result	SE	Control group mean	Estimated interaction	SE	Control group mean
Borrowing									
Number of loans from financial institutions in month before survey	0.094**	0.045	0.359	0.080	0.051	0.385	0.039	0.095	0.244
Number of loans from friends, family, or moneylenders in month before survey	−0.011	0.042	0.286	−0.011	0.045	0.279	0.010	0.104	0.317
Business size									
Number of businesses in household	−0.102*	0.060	1.378	−0.057	0.062	1.354	−0.265	0.181	1.488
Number of paid employees (not including in-kind contributions) in all household businesses	−0.273**	0.123	0.878	−0.214	0.130	0.801	−0.272	0.417	1.220
Subjective well-being									
Life satisfaction (scale: 1–4, 1 = not at all, 4 = very)	0.016	0.063	2.818	−0.024	0.067	2.855	0.209	0.168	2.659
Job stress (scale: −12 to 0: 0 = no stress, −12 = always stressed)	−0.190	0.227	−6.725	0.033	0.254	−6.912	−1.189**	0.513	−5.925
Summary index of above outcomes, optimism, calmness, worry, job satisfaction, decision power, and socioeconomic status	−0.053*	0.030	0.000	−0.043	0.032	−0.014	−0.042	0.082	0.064
			N = 1062–1113						*N*(male) = 160–165

*$P < 0.10$, **$P < 0.05$, ***$P < 0.01$.

We report the intention-to-treat estimate because it more closely maps into the policy parameter of interest: the effect of a credit expansion where the final disposition of the application rests on some discretion by the borrower and/or the loan officer.

After the experiment, we hired researchers from a local university to survey all 1601 applicants in the treatment and control groups. The stated purpose of the survey was to collect information on the financial condition and well-being of micro-entrepreneurs and their households. As detailed below, the surveyors asked questions on business condition, household resources, demographics, assets, household member occupation, consumption, subjective well-being, and political and community participation. Neither the survey firm nor the respondents were informed about the experiment or any association with the lender, so as to avoid potential response bias in the treatment group relative to the control group.

Surveyors completed 1113 follow-up surveys, for a 70% response rate (of the 30% attrition, 81% were not found and 19% refused to be surveyed). Table S1 shows that survey completion was not significantly correlated with treatment assignment, nor with interaction of treatment assignment and baseline covariates in a Wald test.

Ninety-nine percent of the surveys were conducted within 11 to 22 months of the date on which the applicant entered the experiment by applying for a loan and being placed in the pool of marginally creditworthy applicants. The mean number of days between treatment and follow-up was 411 ± 76 (SD).

We then used survey data to measure outcomes Y for estimating intention-to-treat effects with the ordinary least-squares (OLS) specification:

$$Y_i^k = \alpha + \beta^k Assignment_i + \delta Risk_i + \phi APP\_WHEN_i + \gamma SURVEY\_WHEN_i + \varepsilon_i \quad (1)$$

where k indexes different outcomes (e.g., number of formal sector loans in the month before the survey; life satisfaction) for applicant i (or i's household). $Assignment_i = 1$ if the individual was assigned to treatment (regardless of whether they actually received a loan). $Risk_i$ captures the applicant's credit-score window (low or high); the probability of assignment to treatment was conditional on this (set to either 0.60 or 0.85, depending on their credit score), and thus it is necessary to include this as a control variable in all specifications. APP_WHEN is a vector of indicator variables for the month and year in which the applicant entered the experiment; $SURVEY\_WHEN$ is a vector of indicator variables for the month and year in which the survey was completed. These variables control flexibly for the possibility that the lag between application and survey is correlated with both treatment status and outcomes.

Results. The first part of our evaluation focuses on how microcredit is supposed to work in theory, by testing four key hypotheses put forth by microcredit advocates.

H1: Microcredit mitigates market failures. The theory of microcredit focuses on how it alleviates asymmetric information and credit rationing (*29*, *30*). If rationing exists, then there will be excess demand for credit even at market rates (i.e., prices will not clear markets as predicted by canonical neoclassical models). We test whether there was ex ante rationing, and by implication market failure, in our setting by using different types of borrowing as dependent variables in our estimating equation.

Table 2 presents the key results, focusing on "counting" outcomes and a 1-month recall period to minimize noise. We find that FMB's microcredit expansion, at market rates, did significantly increase borrowing from financial institutions. This result is consistent with H1. The point estimate

Table 3. Microcredit in practice: Intention-to-treat effects on household risk management. For the full sample and for females, impacts on trust outcomes are estimated using ordered probit, and other data are OLS results for the independent variable "assigned a loan"; Huber-White SEs and control group means for the dependent variable listed in each row are also shown. The incremental effect on males is shown as an estimate for the interaction between "assigned a loan" and "male." Variation in sample sizes is due to survey question nonresponse. All estimates control for probability of assignment to treatment and for timing of treatment assignment and survey measurement.

	Full sample			Females			Incremental effect on males		
	OLS or ordered probit result	SE	Control group mean	OLS or ordered probit result	SE	Control group mean	Estimated interaction	SE	Control group mean
Financial instruments									
Any health insurance	−0.035	0.038	0.658	−0.018	0.043	0.646	−0.101	0.094	0.707
Any other type of insurance	−0.079**	0.039	0.486	−0.066	0.043	0.475	−0.072	0.101	0.537
Any savings in household	0.002	0.039	0.591	−0.009	0.043	0.594	0.063	0.099	0.575
Family/community networks									
Trust that you would not be taken advantage of (1 = people would take advantage, 10 = people would be fair)	−0.060	0.082	7.685	−0.087	0.092	7.725	0.150	0.192	7.512
Trust in your neighborhood (−4 = no trust, −1 = complete trust)	0.209**	0.090	−2.215	0.203**	0.101	−2.219	0.064	0.196	−2.195
Trust in people you know personally (−4 = no trust, −1 = complete trust)	0.036	0.093	−1.895	0.001	0.102	−1.882	0.215	0.237	−1.951
Trust in your business associates (−4 = no trust, −1 = complete trust)	0.101	0.089	−2.184	0.080	0.101	−2.175	0.117	0.186	−2.225
Could get financial assistance from family or friends in an emergency	0.010	0.027	0.883	−0.003	0.030	0.888	0.080	0.068	0.861
Could get 10,000 pesos' worth of financial assistance from family or friends in an emergency	0.102***	0.040	0.370	0.091**	0.044	0.379	0.062	0.102	0.333
Could get unlimited financial assistance from family or friends in an emergency	0.090**	0.035	0.254	0.074*	0.039	0.267	0.091	0.087	0.194
		N = 995–1113				N(male) = 151–165			

*P < 0.10, **P < 0.05, ***P < 0.01.

implies an economically large effect: a 9.4 percentage point (26%) increase in the proportion of individuals with formal loans. We find no effect on "informal" borrowing: loans from friends, family, and moneylenders (i.e., no crowding out of other debt). Results for other measures of borrowing, and for a 1-year recall period, paint a similar picture (one interesting exception is that we do find a significant treatment effect reducing the likelihood of any informal borrowing over the past year).

The size of the treatment effects may be understated here for two reasons: (i) underreporting of borrowing [e.g., more than half of borrowers known, from FMB's data, to have a loan outstanding from FMB in the month before the survey do not report any borrowing in the survey; see (*31*) for a similar result from a different setting]; and (ii) the timing of the recall period, which may miss earlier (short-run) effects, particularly for borrowers who needed/took only one loan, because the follow-up survey took place 11 to 22 months after treatment assignment, and the initial loan term was only 4 months.

H2: Microcredit spurs business growth. We focus here on two key proxies for the scale of business activity undertaken by our subjects' households. Both are relatively easy to measure precisely. One is the number of businesses the household is operating; the other is the number of paid employees. The results suggest that expanded access to credit shrinks business scale if anything: Our treatment group business owners operate 0.1 fewer business (7% less than the control group mean; $P = 0.09$, t test) and have 0.27 (31%) fewer paid employees. These results do not support H2. Results for three other proxies for business size and success—total profits, gross sales, and inventory—are noisy and hence do not sharpen inference with respect to H2 (table S4).

H3: Microcredit improves subjective well-being. Microcredit proponents argue that having additional choices makes borrowers feel more capable, optimistic, and happy. We use several measures of subjective well-being to test these hypotheses. Table 2 shows that we find a precisely estimated null effect on life satisfaction (*32*). Nor do we find any change in stress in the full sample, although, as in Fernald *et al*. (*33*), we do find a significant (and significantly greater) increase in stress for male borrowers. Combining these and several other standard measures into a summary index of subjective well-being (*12*), we find that expanded credit access produces a small (1/20 of a standard deviation; $P = 0.08$, t test) decrease in subjective well-being. In short, we find no support for H3. One important question, beyond the scope of the present study, is whether (subjective well-being) benefits are specific to the lending mechanism; for example, perhaps group lending produces some benefits that individual lending does not (or vice versa).

H4: There are disproportionate benefits from targeting women. We test this, for each outcome discussed above, by interacting the treatment assignment and gender. We find little evidence of significantly different treatment effects for women (Table 2); the one exception is that males experience a greater increase in stress. Two important caveats are that we lack the statistical power to detect some economically meaningful differences, and that only 15% of our sample is male (raising questions about external validity).

The second part of our evaluation presents some evidence on how microcredit works in practice, by testing two additional hypotheses suggested by the literature on risk sharing (*8*, *34*). Results appear in Table 3.

H5: Microcredit is a substitute for insurance and precautionary saving. We test this by looking at three common ways that households mitigate risk. We find a 3.5 percentage point reduction ($P = 0.36$, t test) for the treatment effect on holding any health insurance, and a 7.9 percentage point reduction, significant at 5% ($P = 0.043$, t test), in holding other types of insurance (life, home, property, fire, and car). These results suggest that microcredit allows households to conserve resources by substituting away from formal insurance to other methods of managing risk. We find no effect on the likelihood of holding savings, with the caveat that there are motives besides precautionary ones for saving. In all, we find limited support for H5, in that access to credit led individuals to lower their demand for formal insurance. This suggests that microcredit is used to buffer fluctuations in income and expenses.

H6: Microcredit improves informal risk-sharing. In theory, this can go either way: Microcredit may disrupt informal arrangements because outside options weaken commitments to reciprocate, or it may strengthen informal arrangements by providing participants with greater resources and liquidity. We find evidence for the latter effect (i.e., for H6). Trust in one's neighborhood increases for the treatment group, as do two other trust measures (*35*). The treatment group also reports greater access to informal credit (as measured by responses to questions about how much financial assistance one could get from friends or family in an emergency).

In all, these results suggest that microcredit improves the ability of household to manage risk by giving them additional options: using credit instead of insurance (or precautionary saving), and strengthening family and community risk-sharing.

Discussion. We used the random assignment of loan offers via credit scoring to generate evidence on the effects of individual liability microcredit on micro-entrepreneurs in Manila. The results are surprising. Entrepreneurs who were randomly assigned microloans shrank their businesses, and their subjective well-being did not improve. The benefits we found—that microcredit access improves risk management—are on margins often deemed second-order by policymakers, practitioners, and economists. We did not find any evidence that treatment effects are more pronounced for female borrowers.

The overall picture of our results questions the wisdom of assuming that impacts are stronger for preexisting micro-entrepreneurs and women than for "consumers," men, or aspiring micro-entrepreneurs. Money is fungible, and we find that entrepreneurs do not necessarily invest loan proceeds in growing their businesses. Perhaps micro-entrepreneurs only increase business investment if loan proceeds are tied to a more detailed business planning exercise or are followed by unusually close monitoring from the lender. In any case, limiting microcredit access to entrepreneurs may forgo opportunities to improve human capital and risk-sharing for non–micro-entrepreneurs (*37*).

Above all, our results highlight the importance of replicating tests of theories and interventions across different settings. Our findings add to a very muddled picture on the impacts (or lack thereof) of microcredit. At this point it remains to be seen whether different studies arrive at different estimates because of true underlying heterogeneity across settings, to different targeting [e.g., Banerjee *et al*. (*11*) found positive impacts on investment in new micro-enterprises but no growth of preexisting micro-enterprises], and/or to differences (and flaws) in some methodologies. One approach to solving this puzzle is to replicate research designs across settings.

Random assignment via credit scoring is a viable tool for replication, as it provides a win-win for lenders looking for an effective way to improve operations, and for other constituencies (researchers, donors, investors, and policymakers) looking for an effective way to measure impacts of expanding access to microcredit. It also provides a way to maximize statistical power given budget and/or operational constraints; the leading alternative randomized method, randomized program placement, will typically require much larger sample sizes because of clustering issues and lower take-up rates (*38*).

Furthermore, regarding statistical power, this study and other similar studies are hindered by low differential take-up in treatment versus control, as well as limited resources for follow-up data collection. Given the results of this study and other similar studies, we suggest that further studies would benefit from larger sample sizes, from greater focus on operational issues that could expand the differential take-up rate of treatment versus control, and perhaps from more waves of follow-up data, in order to expand the set of hypotheses that can be tested with sufficient power.

Nonetheless, random assignment via credit scoring does have its methodological limitations for measuring impacts of microcredit. We highlight five additional issues that researchers might consider when designing future studies.

First, theory suggests that spillovers, both positive and negative, may occur as a result of expanding access to credit. For example, expanding access to credit to some firms may harm their competitors and may help their suppliers or clients.

Measuring such spillovers requires a larger sample size than used here, and an enriched experimental design; for example, one might use two-stage randomization, first stratifying by geographic area or industry and then randomizing the proportion of each industry or geographic area that gets assigned to treatment. Second, many microlenders delegate credit decisions to borrowing groups. Credit scoring (with a random component) could, in principle at least, be provided to groups as an additional tool for use in deciding who gets credit and who does not. Third, new methods for mitigating underreporting of household and business debts may help researchers improve statistical power for estimating impacts of credit expansion on overall borrowing and more ultimate impacts. Fourth, having a loan application rejected may change future behavior (e.g., by discouraging pursuit of credit from another source). It is thus important to take care—as we did in the implementation studied here—to avoid changing the beliefs of rejected applicants (e.g., by rejecting people who are clearly creditworthy), lest the assignment to control (rejected) status induce behavior that differs from what people would have done in the absence of the experiment (*39*). Fifth, different methodologies may produce subject pools. For example, a branch expansion or clear loosening of credit criteria may draw in applicants previously discouraged from applying, in contrast to the methodology here, where the exact (weighting of) criteria that would affect a marginal application decision were opaque both before and after the credit-scoring intervention. Hence, our methodology identifies impacts on marginal borrowers only and may lack external validity for making inferences about inframarginal borrowers (who are clearly above or below the creditworthiness bar prior to the intervention) (*40*). This concern is mitigated in the present study because our cooperating lender deemed 74% of its first-time applicants "marginal." From a policy perspective, marginal borrowers are often the subject pool of interest, even if the pool is smaller (proportionally) than it is here. If inframarginal borrowers are of interest as well, one practical methodological approach might be to introduce some randomness into the maximum loan size offered by the lender.

We conclude by recapping our two key points. One is that the theory and practice of microcredit remain far ahead of the evidentiary base needed to make good policy and to improve the delivery of financial intermediation. Our findings here are surprising and provocative: Microcredit in this context did not generate bigger businesses, higher income, and higher subjective well-being, but rather led to stronger risk management, fewer businesses, and lower subjective well-being. The current literature, and popular rhetoric from policymakers and microcredit institutions, puts forward a relatively simple story about microcredit working through business investment and female empowerment. In contrast, our findings suggest that microcredit works through more complex and disparate mechanisms that start with the household rather than with the business. This leads us to our second point: The route to a clearer picture of whether and how microcredit works is methodological. Replicating sound research designs across different settings will help reveal whether treatment effects truly differ across settings, or whether the existing muddle is due more to methodological differences (and flaws). We hope we have convinced readers that the design used in this study is replicable, and we look forward to continued work with other lenders and researchers to build a better evidentiary base for microcredit policy and practice.

References and Notes

1. Note that whereas microcredit targets small businesses, the individuals may in practice use proceeds for household expenses, nonbusiness investments such as education or health, remittances, etc.
2. B. Armendáriz, J. Morduch, *The Economics of Microfinance* (MIT Press, Cambridge, MA, 2007), pp. 199–230.
3. J. Conning, C. Udry, in *The Handbook of Agricultural Economics, Vol. 3*, R. E. Evenson, P. Pingali, T. P. Schultz, Eds. (Elsevier Science, Amsterdam, 2005), pp. 2857–2908.
4. J. Zinman, *J. Bank. Finance* **34**, 546 (2010).
5. B. Armendáriz, J. Morduch, *The Economics of Microfinance* (MIT Press, Cambridge, MA, ed. 2, 2010), pp. 179–197.
6. R. Cull, A. Demirguç-Kunt, J. Morduch, *Econ. J.* **117**, F107 (2007).
7. R. Cull, A. Demirguç-Kunt, J. Morduch, *J. Econ. Perspect.* **23**, 167 (2009).
8. D. Karlan, J. Morduch, in *Handbook of Development Economics, Vol. 5*, D. Rodrik, M. Rosenzweig, Eds. (Elsevier, Amsterdam, 2009), pp. 4704–4784.
9. See (*8*) for additional details. Also note that the definition of "microcredit" is often debated, but typically includes loans to micro-entrepreneurs that are small but sufficiently large to provide meaningful support to small vendors, convenience stores or production facilities. Standard definitions often exclude "consumer" loans made to salaried individuals.
10. Banerjee *et al.* (*11*) is the first randomized evaluation of group-liability micro-entrepreneurial microcredit, and Karlan and Zinman (*12*) reports on the first randomized evaluation of consumer lending. The key difference between Karlan and Zinman (*12*) and this paper is that the lender here targets entrepreneurs whereas the South African lender in the prior study was closer to a payday lender, making loans to employed individuals without any questions or examination of how funds would be used.
11. A. Banerjee, E. Duflo, R. Glennerster, C. Kinnan, *The Miracle of Microfinance? Evidence From a Randomized Evaluation* (working paper). Massachusetts Institute of Technology, 2010 (http://econ-www.mit.edu/files/4162).
12. D. Karlan, J. Zinman, *Rev. Financ. Stud.* **23**, 433 (2010).
13. The USAID contractor is Chemonics' Microenterprise Access to Banking Services (MABS). MABS provides technical assistance on bank operations, including advice on how to set prices in accordance with normal (i.e., not subsidized) consumer banking practices. MABS provides neither subsidized credit nor incentives to lower interest rates.
14. Opportunity sets refer to investment opportunities, broadly defined. For example, a concern with nonrandomized studies is that those who get credit (because of relatively high demand or high supply) get it because they have businesses or other investment options with relatively high returns, and those who seem similar on observable characteristics but do not choose to borrow simply do not have the same set of business or investment options. A methodology that does not account for, or remove, this sort of underlying correlation may mistakenly attribute causal impacts to microcredit use/ access that are actually due to underlying differences between borrowers and nonborrowers that have nothing to do with microcredit per se. The same concern holds for other characteristics that are difficult to observe (and hence control for) that might be correlated with both outcomes (e.g., business success) and borrowing or lending decisions. The other examples in the text refer to what economists think of as "preferences" (e.g., patience or attitudes toward risk) and "aptitude" (e.g., entrepreneurial skills or business acumen).
15. See (*8*) for a detailed discussion of the field experiments cited above and several other important microcredit evaluations using various methodologies (*16*–*23*)
16. J. Morduch, "Does microfinance really help the poor? New evidence on flagship programs in Bangladesh" (working paper, 1998).
17. M. Pitt, S. Khandker, *J. Polit. Econ.* **106**, 958 (1998).
18. B. Coleman, *J. Dev. Econ.* **60**, 105 (1999).
19. S.-M. McKernan, *Rev. Econ. Stat.* **84**, 93 (2002).
20. M. Pitt, S. Khandker, O. H. Chowdhury, D. Millimet, *Int. Econ. Rev.* **44**, 87 (2003).
21. J. Kaboski, R. Townsend, *J. Eur. Econ. Assoc.* **3**, 1 (2005).
22. J. Kaboski, R. Townsend, "The impacts of credit on village economies" (working paper, 2009).
23. D. Roodman, J. Morduch, "The impact of microcredit on the poor in Bangladesh: Revisiting the evidence" (working paper, 2010).
24. In *Benchmarking Asian Microfinance 2005*, the Microfinance Information eXchange (MIX) reports that Filipino microlenders have the lowest outreach in the region—a median of 10,000 borrowers per MFI (www.themix.org/sites/default/files/MIX_2005_Asia_Benchmarking_Report_EN.pdf).
25. The lender also requires first-time borrowers to open a savings account and maintain a minimum balance of 200 pesos. The real and nominal interest rates are roughly equivalent, as inflation in the Philippines during the study period was in the single digits.
26. One branch was removed from the study when it was discovered that one account officer had found the underlying files saved by the credit-scoring software and altered both the assignment to treatment and data recorded from the application. This was discovered in audits of the proportion assigned to treatment, as well as audits to verify that the handwritten application from the client matched the data entered into the credit-scoring software. No other branches had problems revealed by such audits.
27. "Business capacity" consists of business net cash flow, business assets, and number of employees. "Personal financial resources" consists of personal savings and credit history. "Outside financial resources" considers spousal employment and the value of household assets. "Stability" considers the ownership status and age of the business.
28. Survey completion, is, unsurprisingly, correlated with baseline characteristics. Coupled with orthogonality to treatment assignment interacted with those same baseline characteristics, these results suggest that our results have internal validity but may lack external validity for those who did not complete surveys. See table S1 for details.
29. B. Armendariz de Aghion, J. Morduch, *The Economics of Microfinance* (MIT Press, Cambridge, MA, ed. 2, 2010), pp. 1–24.
30. Whereas canonical microcredit has used group liability mechanisms to try to overcome asymmetric information problems, other efforts, such as the one studied here, use individual liability coupled with attempts to use information that is available to the lender more efficiently.
31. D. Karlan, J. Zinman, *J. Eur. Econ. Assoc. Pap. Proc.* **6**, 510 (2008).
32. A. Oswald, S. Wu, *Science* **327**, 576 (2010); 10.1126/science.1180606.
33. L. Fernald, R. Hamad, D. Karlan, E. Ozer, J. Zinman, *BMC Pub. Health* **8**, 409 (2010).
34. X. Giné, *J. Dev. Econ.* (2010).
35. See Cleary and Stokes (*36*) for more detail on the trust measures.
36. M. R. Cleary, S. C. Stokes, *Democracy and the Culture of Skepticism: Political Trust in Argentina and Mexico* (Russell Sage Foundation, New York, 2006).
37. Karlan and Zinman (*12*) find direct evidence that salaried workers benefit from microloans.

38. However, as was the case in this study, operational considerations may dictate less than optimal experimental design regarding the proportion allocated to treatment versus control. Under normal circumstances and equal expected variance in treatment and control, power is optimized when treatment and control groups are equalized; however, the allocation rule in this study was not evenly split treatment and control groups because of portfolio management constraints.
39. Concerns about John Henry effects would affect estimation strategy: Treatment-on-the-treated estimates, using instrumental variables, would lead to biased estimates. Intent-to-treat estimates remain unbiased but make comparisons across settings more difficult if treatment compliance/take-up rates are quite heterogeneous.
40. Those with lower credit scores in our sample may be closer to the types of inframarginal applicants who might be encouraged to apply under a more visible expansion or loosening of credit. Tables S2 and S3 show little evidence of differential treatment effects for those with lower versus higher credit scores in our sample.

Acknowledgments: We thank J. Bauchet, L. Crowley, D. Duthie, M. Duthie, E. Ganir, K. Haggag, T. Harigaya, J. Soriano, M. Startz, and R. Zarsuelo for outstanding project management and research assistance; N. Hite, D. McKenzie, D. Roodman, and seminar participants at the Center for Global Development and many academic seminars and conferences for helpful comments; Bill and Melinda Gates Foundation and the National Science Foundation for funding; and especially J. Owens and his team at the USAID-funded MABS program for help with the project. Any views expressed are those of the authors and do not necessarily represent those of the funders, MABS, or USAID. Above all we thank First Macro Bank for generously providing the data from its credit-scoring experiment.

Supporting Online Material
www.sciencemag.org/cgi/content/full/332/6035/1278/DC1
Tables S1 to S4
Data set and do files

8 November 2010; accepted 3 May 2011
10.1126/science.1200138

Local Macrophage Proliferation, Rather than Recruitment from the Blood, Is a Signature of T_H2 Inflammation

Stephen J. Jenkins,[1] Dominik Ruckerl,[1] Peter C. Cook,[1] Lucy H. Jones,[1] Fred D. Finkelman,[2–5] Nico van Rooijen,[6] Andrew S. MacDonald,[1] Judith E. Allen[1*]

A defining feature of inflammation is the accumulation of innate immune cells in the tissue that are thought to be recruited from the blood. We reveal that a distinct process exists in which tissue macrophages undergo rapid in situ proliferation in order to increase population density. This inflammatory mechanism occurred during T helper 2 (T_H2)–related pathologies under the control of the archetypal T_H2 cytokine interleukin-4 (IL-4) and was a fundamental component of T_H2 inflammation because exogenous IL-4 was sufficient to drive accumulation of tissue macrophages through self-renewal. Thus, expansion of innate cells necessary for pathogen control or wound repair can occur without recruitment of potentially tissue-destructive inflammatory cells.

Recruitment of leukocytes from the blood is the key feature of inflammatory responses to tissue damage or infection. Prominent in this cascade is the rapid influx of granulocytes and monocytes that subsequently differentiate into inflammatory macrophages and/or dendritic cells (*1*, *2*). This paradigm of classical "type 1" inflammation is firmly grounded in processes elicited by microbial infection or necrotic cell death (*2*). Inflammation driven by helminth infection or allergy represents a distinct challenge to the immune system and is characterized by the recruitment of cells that produce the cytokine interleukin-4 (IL-4), including eosinophils, basophils, and $CD4^+$ T helper 2 (T_H2) cells (*3*, *4*). Like classical inflammation, "type 2" inflammation also results in the accumulation of large numbers of macrophages in the affected tissues (*3*, *5*–*7*), yet despite this, IL-4 and T_H2 responses are often regarded as "anti-inflammatory."

A key distinguishing feature of macrophages in type 2 inflammation is their polarization toward an alternative (also known as M2) state of activation, mainly driven by IL-4 and IL-13 and characterized by expression of a distinct repertoire of molecules including arginase I, resistin-like molecule alpha (RELMα), and Ym1/2 (*8*). It has been widely assumed (*1*, *8*) that accumulation of alternatively activated macrophages occurs through a process of recruitment similar to macrophage influx during classical inflammation. Supporting this view, depletion of blood monocytes impairs recruitment of mucosal M2 macrophages during gastrointestinal nematode infection (*5*). The intestinal environment presents unique challenges for the study of type 2–driven inflammation, however, because commensal bacteria may act as a classical trigger for monocyte influx.

M2 macrophages accumulate independently of monocytes. To examine type 2 inflammatory processes, we profiled cell recruitment during infection with a nematode that resides within "sterile" tissues (*9*). The rodent filarial nematode *Litomosoides sigmodontis* induces strongly T_H2-biased responses (fig. S1A) and drives alternative macrophage activation in the pleural cavity (Fig. 1A), after migrating there from the skin (Fig. 1B). For comparison, intrathoracic injection of thioglycollate was used as a benchmark of the classical inflammatory cascade (*10*), because when injected, it induces recruitment of macrophages to the pleural cavity. Indeed, thioglycollate injection resulted in the rapid influx of neutrophils ($Gr\text{-}1^{high}/Ly\text{-}6G^{high}$) and $Gr\text{-}1^{intermediate\ (int)}/Ly\text{-}6C^{high}$ monocytes, leading to accumulation of large numbers of $F4/80^+$ macrophages by day 3 (Fig. 1C and fig. S1B). In contrast, *L. sigmodontis* infection triggered little neutrophil or $Gr\text{-}1^{int}/Ly\text{-}6C^{high}$ monocyte recruitment to the pleural cavity during the first 15 days after infection (Fig. 1D and fig. S1C), even though larvae arrive at this site between 3 and 6 days after infection (Fig. 1B). Despite this, large numbers of $F4/80^+$ macrophages gradually accumulated in the cavity after day 6 of infection (Fig. 1D), accompanied by their conversion to an alternatively activated phenotype (Fig. 1A). There was also a marked difference in surface phenotype of the macrophage populations elicited by nematode infection compared to classical inflammation: Those elicited by infection expressed levels of F4/80 similar to that of resident $F4/80^{high}$ macrophages from naïve tissues (fig. S1C), whereas those induced by thioglycollate expressed low F4/80 levels (fig. S1B), which is characteristic of macrophages recruited to the serous cavities under inflammatory conditions (*11*).

To determine the role of monocytes in macrophage accumulation, we injected clodronate-loaded (CL) liposomes intravenously (iv). CL-liposomes block tissue infiltration by macrophages in a variety of inflammatory settings (*5*, *12*, *13*). This procedure depletes both $Gr\text{-}1^{int}/Ly\text{-}6C^{high}$ monocytes, which are the precursors of $F4/80^{low}$ inflammatory macrophages recruited during classical inflammation, and $Gr\text{-}1^-/Ly\text{-}6C^-$ monocytes, which have been proposed as precursors of alternatively activated and tissue-resident macrophages (*1*, *14*). As expected, treatment with CL-liposomes blocked the accumulation of macrophages in the pleural cavity induced by thioglycollate injection (Fig. 1E). In contrast, administration of CL-liposomes during the period when macrophages accumulate in the cavity (Fig. 1F), or when worms first enter the cavity (fig. S2), had no effect on either the number of pleural macrophages at day 10 after *L. sigmodontis* infection or the frequency of those expressing alternative activation markers. Macrophage accumulation occurred despite almost complete removal of $Gr\text{-}1^-/Ly\text{-}6C^-$ monocytes from the blood

[1]Centre for Immunity, Infection and Evolution, and the Institute for Immunology and Infection Research, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JT, UK. [2]Department of Medicine, Cincinnati Veterans Affairs Medical Center, Cincinnati, OH 45220, USA. [3]Division of Immunology, Allergy, and Rheumatology, Department of Internal Medicine, University of Cincinnati College of Medicine, Cincinnati, OH 45267, USA. [4]Division of Immunobiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA. [5]Division of Rheumatology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA. [6]Department of Molecular and Cell Biology, Free University Medical Centre, Amsterdam, Netherlands.

*To whom correspondence should be addressed. E-mail: j.allen@ed.ac.uk