ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Haushofer, Johannes; Mudida, Robert; Shapiro, Jeremy

Working Paper

The Comparative Impact of Cash Transfers and a Psychotherapy Program on Psychological and Economic Well-being

IFN Working Paper, No. 1377

Provided in Cooperation with:
Research Institute of Industrial Economics (IFN), Stockholm

Suggested Citation: Haushofer, Johannes; Mudida, Robert; Shapiro, Jeremy (2021) : The Comparative Impact of Cash Transfers and a Psychotherapy Program on Psychological and Economic Well-being, IFN Working Paper, No. 1377, Research Institute of Industrial Economics (IFN), Stockholm

This Version is available at:
https://hdl.handle.net/10419/240520

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

IFN Working Paper No. 1377, 2021

The Comparative Impact of Cash Transfers and a Psychotherapy Program on Psychological and Economic Well-being

Johannes Haushofer, Robert Mudida and Jeremy Shapiro

Research Institute of Industrial Economics
P.O. Box 55665
SE-102 15 Stockholm, Sweden
info@ifn.se
www.ifn.se

The Comparative Impact of Cash Transfers and a Psychotherapy Program on Psychological and Economic Well-being*

Johannes Haushofer† Robert Mudida‡ Jeremy Shapiro§

January 4, 2021

Abstract

We study the economic and psychological effects of a USD 1076 PPP unconditional cash transfer, a five-week psychotherapy program, and the combination of both interventions among 5,756 individuals in rural Kenya. One year after the interventions, cash transfer recipients had higher consumption, asset holdings, and revenue, as well as higher levels of psychological well-being than control households. In contrast, the psychotherapy program had no measurable effects on either psychological or economic outcomes, both for individuals with poor mental health at baseline and others. The effects of the combined treatment are similar to those of the cash transfer alone.
JEL codes: C93, O12, D90

*We thank the participants for their time; Tilman Graff, Magdalena Larreboure, Esther Owelle, and Michala Riis-Vestergaard for excellent research assistance; the staff of the Busara Center for Behavioral Economics; and the staff of the NGO delivering the intervention, for fruitful collaboration and comments. We thank Victoria Baranov, Barbara Biasi, Chris Blattman, Michelle Craske, Julian Jamison, Crick Lund, Joanna Maselko, Petra Moser, Robert Östling, Miranda Wolpert, and seminar and conference participants at the World Bank, BREAD, St. Andrews, LSE, Tilburg, Michigan, University of Pennsylvania, University of Connecticut, Northeastern University, UCLA, and Bocconi, for comments. This study was funded by grants to Jeremy Shapiro by the Gates Foundation and an Anonymous Donor, and by Princeton University. Johannes Haushofer is grateful for financial support from Jan Wallanders och Tom Hedelius stiftelse. IRB approval was obtained from Princeton University (protocol no. 7375) and the Kenya Medical Research Institute (Non-KEMRI protocol no. 531). The study was pre-registered at https://www.socialscienceregistry.org/trials/3928.

†Stockholm University, NBER, BREAD, Busara Center for Behavioral Economics, MPI for Collective Goods, and Research Institute of Industrial Economics. johannes.haushofer@ne.su.se

‡Strathmore Business School. rmudida@strathmore.edu

§Busara Center for Behavioral Economics. jeremy.shapiro@busaracenter.org

Recent work in economics and psychology has suggested that poverty and low psychological well-being may mutually reinforce each other (Ridley et al. 2020; Lund et al. 2011). For example, poverty may lead to stress, which over time can cause mental illness (Hammen 2005; Haushofer and Fehr 2014). Conversely, poor mental health may affect labor market participation and productivity, and thus may exacerbate poverty (Schoenbaum et al. 2002; Biasi, Dahl, and Moser 2019; Ravensteijn and Schachar 2018). This potential feedback loop suggests that interventions that alleviate poverty may be effective in both reducing poverty and improving mental health. Conversely, interventions that target mental health may improve both mental health itself, and poverty. In addition, if mental health and poverty are strong complements, interventions that address both poverty and mental health may be particularly effective in improving both outcomes.

Here we compare the individual and joint impact of an unconditional cash transfer and a psychotherapy program on measures of both economic and psychological well-being. We randomly assigned 5,756 low-income households in rural Kenya to one of four treatment conditions. 540 households in 60 villages received an unconditional cash transfer of USD 1,076 PPP (USD 485 nominal; "CT").[1] 525 households in 60 different villages were treated with Problem Management Plus ("PM+"), a mental health intervention delivered by trained community health workers (CHWs). This program was developed specifically for low-resource settings by the World Health Organization (WHO), and is generally considered a flagship psychotherapy intervention for these contexts. Importantly, it has previously been shown to be effective in improving mental health in Kenya (Bryant et al. 2017).[2] A third group of 493 households in another 53 villages received both cash transfers and PM+ ("CT&PM+"). Comparing outcomes of these groups with a pure control group of 1703 households in another group of 60 villages allows us to determine the causal impact of

[1]The PPP rate for Kenya at the time of the study was 46.49. The transfer corresponds to about 20 months of per capita control group consumption.

[2]Bryant et al. (2017) report a 0.57 standard deviation (SD) improvement in the General Health Questionnaire (GHQ-12) among women who had been exposed to gender-based violence 3 months after they participated in the program as part of a randomized experiment.

these interventions. In villages receiving the CT-only and PM+-only interventions, we also surveyed 1,237 and 1,219 non-recipient households, respectively, allowing us to estimate spillover effects of these interventions at the village level. In CT&PM+ villages and CT villages, we randomly allocated half of cash transfer recipients to receiving a lump sum transfer, and the other half a sequence of 5 weekly transfers, to study the impact of transfer frequency.

We administered a detailed survey both before the interventions, and about a year after they ended. The cash transfer was effective in improving both psychological and economic well-being: we find a 20 percent increase in monthly consumption in cash transfer recipients relative to the control group about a year after the intervention; a 47 percent increase in asset holdings; and a 26 percent increase in household revenue. An index of psychological well-being increases by 0.23 standard deviations (SD), and scores on the GHQ-12 mental health questionnaire—a widely used screening instrument for psychiatric morbidity—improve by 0.16 standard deviations. In contrast, the PM+ intervention affects neither economic nor psychological outcomes after one year: consumption and asset holdings both increase by 5 percent, and revenue by 9 percent, but these effects are not statistically significant. Similarly, the psychological well-being index is not affected by the PM+ treatment, showing a non-significant effect of -0.01 SD. The components of the index, such as psychological distress, happiness, and life satisfaction, also show no treatment effect.[3] Importantly, the PM+ treatment also has no effects on psychological well-being or mental health amongst those participants who had high psychological distress at baseline. We also find no evidence that PM+ has early positive effects that dissipate quickly over time: even those participants who received the program most recently relative to the endline survey (about 7 months prior) show no improvement in psychological well-being at endline.

Outcomes in the group that received both the cash transfer and the PM+ intervention are similar to those in the group that received only the cash

[3]We observe a 0.24 SD increase in intimate partner violence (IPV) reports among female PM+ recipients, but this effect is not significant after adjusting for multiple comparisons, and is not observed in more alternative measures of IPV that minimize response bias.

treatment: asset holdings increase by 41 percent, revenue by 17 percent, and psychological well-being by 0.27 SD. Consumption shows a somewhat smaller treatment effect than in the cash-only group, with a non-significant 7 percent increase, but the 95 percent confidence interval includes increases up to 18 percent. Weekly cash transfers are somewhat more effective in improving economic outcomes compared to lump-sum transfers, and have similar impacts as lump-sum transfers on psychological well-being. We find little evidence of spillovers of either intervention to non-recipient households.

Thus, our main finding is that after one year, cash transfers have larger effects on economic and psychological well-being than the PM+ mental health intervention, which itself produces no improvements relative to control. Importantly, the per-household cost of the cash transfers was lower than that of the PM+ intervention. The PM+ intervention is therefore both more expensive and less effective in affecting our outcomes of interest.[4]

What explains the negligible impacts of PM+, even on outcomes which it specifically targets, such as psychological well-being? First, we were powered to detect PM+ treatment effects of 0.17 SD on psychological well-being. This compares favorably to typical effect sizes of psychotherapy on psychological outcomes, which are around 0.6–0.7 SD in the literature as a whole (Cuijpers et al. 2010; Cuijpers et al. 2013), 0.49 SD across low- and middle income countries (Singla et al. 2017), and 0.57 SD in a previous study of PM+ in Kenya (Bryant et al. 2017). Thus, lack of power is unlikely to explain our findings. To further confirm this claim, we use an approach based on Bayesian statistics which is not commonly employed in field experiments, but which can provide useful information. We ask: What is the probability that the *observed* null effect of PM+ reflects a *true* underlying null effect? This probability is given by the *negative predictive value* (NPV), defined as the share of null observations that reflect true null effects, rather than false negatives. We show that even with a 90 percent prior in favor of its effectiveness, the post-study

[4]If only marginal costs of the PM+ program are considered, i.e. per-session payments to the CHWs without any administrative costs, the PM+ intervention may be more cost-effective for some outcomes if the positive but non-significant point estimates of PM+ are taken seriously.

probability that PM+ produces effects larger than 0.40 SD in our setting is only 10 percent.

Second, we largely rule out experimenter demand effects (i.e. a desire by participants to give responses that conform to the expectations of the surveyors), using an approach proposed by De Quidt, Haushofer, and Roth (2018): when we present participants with explicit hypotheses about what we expect them to respond to self-report questions about depression, their responses do not change measurably relative to when they do not receive this information, suggesting that demand effects do not play an important role in our self-report measures. Note also that in our setting, demand effects would be likely to increase rather than reduce observe treatment effects.

Third, notice that the success of the PM+ intervention in a previous study in Kenya (Bryant et al. 2017) limits the explanatory power of several other explanations. Bryant et al. (2017) delivered PM+ to urban women who had experienced gender-based violence, and found improvements in psychological distress 3 months later. Importantly, the short (5 week) duration of the program; its content; and the fact that it is delivered by laypeople were the same in this previous study, arguing against these variables as explanations for our null effect. In addition, the implementation of the program in our study was done by the same NGO as in this previous study in Kenya. Thus, implementation differences are unlikely to explain the lack of impact.

The previous study differed in four ways from the present one: it was conducted in an urban rather than a rural setting; impacts were measured after 3 months rather than 1 year; it was delivered to women who had experienced gender-based violence, rather than a general population sample; and it employed 23 CHWs to deliver PM+ to 209 participants, while the present study employed 72 CHWs to deliver PM+ to 1018 participants, possibly resulting in a loss of fidelity due to scale.[5] We find it unlikely that the urban setting accounts for the positive effects in the earlier study. As mentioned above, the

[5]In a trial of a layperson-delivered psychotherapy for perinatal depression in South Africa, Lund et al. (2020) attribute their null results to a combination of lack of fidelity of intervention delivery and lack of competence of non-specialist, inexperienced CHW counsellors.

delay between intervention and endline is also unlikely to explain our null results, as we observe no detectable effects even for participants who received the intervention 7 months prior.[6] Heterogeneity analysis in our study shows that PM+ has no effects even for female participants, and for female participants who experienced high levels of intimate partner violence. Finally, we consider it unlikely that the intervention was low-fidelity in our setting, because the same NGO used the same extensive training and supervision protocols they had developed during the previous study.

A possible remaining explanation is that the program in the previous study, even though it was nominally the same as in our study, was delivered with the explicit goal of addressing gender-based violence. It is conceivable that the same program may be more effective when it is deployed to solve a specific problem. This mechanism would also explain the large positive effects of a similar program after seven years found by Baranov et al. (2020): this program was geared to alleviate post-partum depression. Similarly, Ghosal et al. (2016) find an increase in savings and health behavior 15 months after a self-esteem training for sex workers in India. Here, too, the stigma associated with sex work provided a clear target for the intervention. Thus, a likely explanation for the lack of lasting PM+ impacts in this study, despite successes in other studies, is that the program was delivered as a general-purpose intervention without a pre-defined problem to address.

It should be noted that the programs studied by Baranov et al. (2020) and Ghosal et al. (2016) were also more intensive, with 16 and 8 sessions, respectively, compared to 5 in PM+. Intensity alone is unlikely to explain our null effects, because the five sessions of the PM+ program did produce short-term effects in Bryant et al. (2017). But intensity may combine with circumscribed intervention goals to produce lasting effects. Thus, the most successful interventions may be those that are high-intensity and have specific goals.

[6] It remains possible that the PM+ program had positive effects after 3 months, but that they had dissipated after 7 months. However, this would have to be a very steep decline, which we find somewhat implausible.

This explanation is also consistent with the results of several other recent studies which find positive effects of psychological interventions on psychological and economic outcomes: they either target more circumscribed problems, or, when they do not, are either more intensive or measure impacts over a shorter time horizon. For example, McKelway (2020) shows that a nine-session self-efficacy training in India increases women's labor market participation four months later; Bernard et al. (2014) find that a single "aspirations" video increases human capital investment in Ethiopia after six months; and Haushofer, John, and Orkin (2019) show that two short trainings aimed at time preferences and executive function in Kenya increase drinking water chlorination 3 months later.

A caveat to this explanation is the closely related study by Blattman, Jamison, and Sheridan (2017), who use a similar design as ours to study the effects of a USD 200 cash transfer and a psychotherapy intervention in Liberia, both alone and in combination. At short time horizons (several weeks), they find impacts of both interventions on economic and psychological outcomes, including depression and distress. At a time horizon of one year, neither of the individual treatments has effects on economic outcomes or depression and distress, although the combined treatment does. These findings are similar to ours in that we also find that the psychotherapy program has limited effectiveness after one year. However, the psychotherapy intervention in Blattman, Jamison, and Sheridan (2017) was much more intensive than ours, with 3 weekly sessions over 8 weeks; and it aimed to address a specific problem, namely crime and violence. This study therefore suggests that an even intensive intervention that targets a specific problem may not be sufficient on its own to generate lasting psychological and economic effects.[7] It therefore remains an open question precisely how program intensity and specificity combine to create lasting impacts.

Our study also contributes to literatures in economics and psychology that

[7]The fact that Blattman, Jamison, and Sheridan (2017) find little impact of the cash transfer after one year may result from the smaller magnitude of the transfers, or differences in study population and context.

study the effect of cash transfers on economic and psychological well-being. In particular, a number of studies have shown that unconditional cash transfers improve both economic and psychological outcomes in developing countries, such as Blattman et al. (2016) in Uganda and Baird, De Hoop, and Özler (2013) in Malawi.[8] More generally, economic interventions such as asset transfer and ultra-poor graduation programs often have positive impacts on both economic and psychological outcomes (Bandiera et al. 2017; Banerjee et al. 2015; Bedoya et al. 2019). Our study confirms these results, and suggests that the impact of cash transfers is rather robust across studies and settings. In fact, with a few exceptions, the impacts of cash transfers on both economic and psychological outcomes in this study are remarkably similar to those we reported previously for a similar program in a different region in Kenya (Haushofer and Shapiro 2016). Even within the present study, the effects of the cash-only treatment arm and the combined cash and PM+ treatment are remarkably similar across most outcomes. Cash therefore appears to be a robust intervention that reliably improves economic and psychological well-being.

The remainder of this paper is structured as follows. Section I describes the design and methods of the study. Section II summarizes the analytic approach. Section III presents results. Section IV concludes.

I. Design and methods

The design and methods were specified in a pre-analysis plan available at https://www.socialscienceregistry.org/trials/3928 (Haushofer, Riis-Vestergaard, and Shapiro 2019). An overview of the design is shown in Figure 1, and details on timing are given in Online Appendix Table A.1.

[8] See Blattman, Fiala, and Martinez (2020) for evidence on limited long-term effects.

I.A Village selection and village randomization

The study was conducted in partnership with the a large international NGO in Nakuru County in Central Kenya. Nakuru County was selected due to high levels of poverty, high baseline rates of poor mental health (at baseline, 33 percent of our sample were classified as suffering from psychological distress; see Section II.C), and because the NGO already had established an infrastructure there. They provided a list of villages in the region in which they were prepared to work. 233 villages were chosen randomly for the study from this list.[9] A map of the treatment area is shown in Figure 2.

In these 233 villages, we conducted a two-stage randomization: at the village level, and at the household level. In the first stage, each village was assigned to one of four villages types: cash transfers only (CT); PM+ only (PM+); cash transfers and PM+ (CT&PM+); and pure control (PC). For the 53 CT&PM+ villages, this happened before baseline data collection, and for the village types, after baseline data collection, with equal numbers among the remaining 180 villages randomized into each remaining group.[10] This randomization was stratified on dichotomized village-level indices of assets, psychological well-being, and prevalence of IPV, as well as on number of surveys, proportion female, and proportion registered on M-Pesa, a digital payments

[9]The village classifications generally followed Kenyan administrative boundaries. Following internal procedures, the NGO further subdivided larger official village boundaries into smaller units to ease the logistics of intervention delivery. This affected around 25 percent of villages. The villages were delivered on two different lists: one in October 2016, and one in January 2017. The first list contained 202 villages. One village from this list was found to not exist, and one was identified to be overlapping entirely with another village on the list. These villages were excluded, and the remaining 200 villages were retained. The second list contained 100 villages. 37 villages from this list were chosen so as to maximize the geographical distance between them. Among these 237 villages, 4 were randomly selected to be training villages for the community health workers who were to deliver PM+ and excluded from the remainder of the study, leaving us with 233 villages.

[10]The reason for the selection of CT&PM+ villages before baseline data collection was that a different number of households was targeted for surveying in these villages compared to the other village types; we therefore needed to know even before baseline which villages these would be. More detail follows in the next section. One village was assigned to the CT&PM+ treatment group after baseline data collection to ensure a total number of CT&PM+ treatment recipients of approximately 500. This village was selected at random before the remaining randomization was conducted.

platform operated by the mobile provider Safaricom and commonly used in Kenya.

I.B Household selection and household randomization

After village selection and randomization of the 53 villages to the CT&PM+ condition, we conducted a baseline survey. Our target was to survey 10 households in each of the 53 CT&PM+ villages, which would later all receive treatment; and 30 households in each of the other villages. In the PM+ and CT villages, these 30 households would later be randomized into 10 treatment and 20 control ("spillover") households; in the PC villages, all 30 would be control households. As mentioned above, the different number of households targeted during baseline in the CT&PM+ villages relative to the other village types explains why these villages were chosen before baseline data collection.

Household selection within villages proceeded as follows. A community entry meeting was conducted with village elders, in which they were provided information about the study. A team of field officers (FOs) was then sent to each village (3 FOs per village for the CT&PM+ villages, 10 per village for the other village types). With the help of local guides assigned by the village elder, the FOs drew up a list of households which met our targeting criterion, which consisted of living in a house without brick, stone, or metal walls. They then randomly chose 10 of these households. In all villages except CT&PM+ villages, the FO additionally selected the two eligible households that were geographically closest to each initial household, for a total of 30 households per village. This geographical approach to sampling ensured that treatment and spillover households in each village were in close geographical proximity to each other. One person in each selected household was chosen as the respondent and surveyed. If only one person was home, this person was surveyed, and if more than one person was home the FO surveyed the household head, or if the household head was not available, whoever greeted the FO at the door (unless they insisted that another household member was surveyed).[11] The survey

[11] For 20 percent of households, respondents were coded as belonging to different villages

respondent in each household later became the recipient of the intervention if the household was randomized into treatment. Thus, household, respondent, and recipient refer to the same person (except for the IPV survey, detailed below). The FOs were unaware of the village type assigned to each village, and baseline data is therefore independent of treatment status.

The respondents surveyed in the baseline survey were then considered for randomization as follows. We first removed 10 percent of all respondents, stratified by village, to be used as the treatment group in another study on digital financial services (DFS). This study tested the effects of assisting respondents with setting up an M-Pesa account, relative to no intervention. Thus, only respondents who were not already registered on M-Pesa, 28 percent of the entire sample, were eligible for the DFS study. In total, 514 respondents were removed from the 233 study villages in this way (523 from the initial sample of 237 villages that includes the four PM+ training villages). Among the remaining respondents, 21 percent were not registered on M-Pesa at baseline; these households serve as the control group for the DFS study, and form part of the control group for the present study. In light of this design, we overweight respondents without M-Pesa accounts in the analysis; details are given below.

Among the remaining households, we randomly assigned one third in each of the CT and PM+ villages to "treatment" status, and two thirds to "spillover" status. This randomization was done within village, and independently of whether households were chosen with the first 10 households in a village, or as part of the 20 neighboring households. The randomization was stratified on dichotomized individual-level indices of assets, psychological well-being, and prevalence of IPV, as well as on dummies for whether the respondent was female and whether he/she was registered on M-Pesa. The 10 households in each CT&PM+ village were all assigned to treatment, and the 30 households in each PC village were all assigned to control.

Finally, all CT and CT&PM+ households were randomized into receiving

by village elders vs. the NGO's data. We resolved such cases by assigning such respondents to the village which was closest to their GPS location. Because this decision was determined solely by GPS coordinates, it was independent of any other variables, including treatment status.

the CT as either a lump-sum transfer of KES 50,000, or in installments of KES 10,000 each week for five weeks to match the delivery of PM+. This randomization was done at the village level, but without further stratification.

I.C Interventions

The interventions were delivered between May 2017 and January 2018 (with one month's pause in August 2017 due to the Kenyan national elections, and two weeks' pause in October/November 2017 due to re-election after the first election results were contested). In each treatment household, the person targeted for the intervention was the main survey respondent, chosen as described above.

I.C.1 Problem Management Plus (PM+)

Intervention structure PM+ is a psychotherapy intervention based on cognitive behavioral therapy (CBT), developed by the WHO for resource-poor contexts (World Health Organization 2018). It was previously adapted to the Kenyan context, and a previous randomized experiment in Kenya confirmed that it reduced psychological distress (Bryant et al. 2017). The intervention consists of five 90-minute one-on-one sessions over five weeks, with one session per week delivered by a CHW. PM+ aims to help the participants reduce problems that they identify as being of concern to them. The program teaches five strategies aimed at reducing psychological distress and improving psychological well-being: problem-solving, managing stress, managing problems, behavioral activation ("get going, keep doing"), and strengthening social support. Each session has one primary focus: Session 1 orients the client to the intervention so as to improve engagement, provides education about common reactions to adversity, and teaches the participant a simple stress management strategy consisting of breathing exercises. Session 2 addresses a participant-selected problem through problem-solving techniques, and introduces behavioral activation. Sessions 3 and 4 continue to support participants' application of problem-solving, behavioral activation, and relaxation exercises, and intro-

duce strategies to strengthen social support. Session 5 reviews the learned strategies, provides education about retaining treatment gains, and the intervention ends. Each session concludes with 10–20 minutes of the "managing stress" exercise.

Training and supervision The delivery of PM+ was run by a large international NGO. They hired and trained 2 clinical supervisors and 10 community health assistants (CHAs), who in turn hired and trained 72 community health workers. The clinical supervisors were psychologists with Bachelors degrees in Psychology. The CHAs were level-one staff in Kenya's health care system, paid by the county government, and trained in medical schools as nutritionists, pharmacists, public health officers, nurses, or lab technicians.

The CHWs delivered the PM+ intervention to respondents. They were volunteers who on average had completed secondary school; their typical occupations were farming, business, teaching, and casual work. Each CHW received 9 full days of classroom training by clinical supervisors or CHAs. In the first training, the 2 clinical supervisors trained 38 CHWs. In the second training, the CHAs trained 34 CHWs under close supervision by the clinical supervisors. Classroom training included: information about common mental health problems (i.e. depression, anxiety, stress); training on the strategies employed by PM+ for treating psychological distress; basic helping skills; and role-play. It followed a manual provided by the WHO (World Health Organization 2018). Following classroom training, each CHW saw at least one client for five fully supervised training sessions. All of these sessions were attended by both a clinical supervisor and CHAs. Because there were 72 CHWs, there were $72 \times 5 = 360$ such in-field supervised training sessions. These "training clients" were not included in our study. CHWs were paid KES 700 for each session of PM+ delivered to one respondent.

During intervention delivery, quality monitoring was done through supervisions and fidelity checks. Supervisions were structured as follows: CHWs met with a clinical supervisor or a CHA for about 2 hours in groups of 4–12 to discuss progress and challenges. 25 percent of the supervisions that were

run by a CHA were also attended by a clinical supervisor to ensure quality. In the first three months after PM+ training, CHWs participated in weekly supervisions. In the following three months, supervisions took place biweekly, followed by monthly supervision in the last three months. Clinical supervisors also provided supervisions to the CHAs after each cohort (see next section for the definition of cohort), and on occasion when supervision was needed at other times. They were also available to be consulted in difficult cases, e.g. when a CHW faced difficulties in the field that surpassed their ability to assist, such as imminent risk of suicide.

Fidelity checks were structured as follows: The clinical supervisors randomly attended PM+ sessions 2, 3, and 4 to perform fidelity checks, i.e. ensure the PM+ guidelines were followed by the CHWs. Each CHW received one such check during each cohort, on rare occasions more than one. During fidelity checks, a score was assigned to the CHW; if the score was low, the supervisor followed up with the CHW, and if no improvement was found over several rounds, the CHW's contract was terminated. A 2–3 day refresher training was conducted by the clinical supervisors, with the assistance of the CHAs, for all CHWs after each cohort, with the length determined by the need observed during supervisions and fidelity checks.

Intervention delivery and compliance The program was rolled out in cohorts, constrained by the number of CHWs available at a given time in each of the two sub-locations of Nakuru County in which the study took place (Naivasha and GilGil). Each CHW treated 3–4 respondents at a time for a total of five weeks (one session per respondent per week; 3–4 sessions per CHW per week). All treatment households within a village were treated at the same time, and the order of villages was randomized. PM+ respondents were contacted either by phone or at home by the CHW and informed about the details of the PM+ program. If the respondent consented to enter the program, the CHW either continued immediately to Session 1 (if contact was in-person and the respondent was available), or scheduled a start-date. Participation in the PM+ intervention was voluntary, and participants were free to withdraw

at any time.

According to the NGO's data, of the 1018 participants randomized into the PM+ treatment across the PM+ and CT&PM+ treatment arms, 968 (95.1 percent) attended all five sessions of the program. 3 (0.3 percent) attended 3 sessions, 8 (0.8 percent) attended one session, and 39 (3.8 percent) attended no sessions. All of these participants are considered treated for analysis purposes (intent-to-treat).

We independently verified these attendance numbers 18 months later by calling recipients of the PM+ and CT&PM+ treatments and asking them to recall the intervention. We tried to call all 525 PM+ and all 493 CT&PM+ recipients, and reached 369 PM+ (70 percent) and 400 CT&PM+ participants (81 percent). Among these, 90 percent of PM+ and 87 percent of CT&PM+ recipients confirmed they received some PM+ sessions. Our phone resurvey could not confirm the high rates of participants receiving the entire schedule, with only 35 percent (20 percent) of PM+ (CT&PM+) subjects remembering having received all five sessions. The median reported number of sessions received was 4 and 3 for PM+ and CT&PM+ recipients, respectively.

I.C.2 Cash transfers (CT)

Cash transfers were delivered and overseen by Busara Center for Behavioral Economics. In the CT&PM+ treatment group, cash transfers were made one month after the beginning of the PM+ program on average (Online Appendix Table A.1 and Figure A.1). Before sending the respondents a cash transfer, they were contacted over the phone or in person by a Busara field officer, and informed that they had been entered into a lottery and their name had been selected to receive KES 50,000 (USD 485 nominal, USD 1076 PPP at the time of the study). This transfer corresponded to about 20 months of per capita consumption (using the monthly per capita control group consumption of USD 52.49 PPP at endline; we did not collect detailed consumption data at baseline). The field officer emphasized that the cash transfer was entirely unconditional ("The money is yours to do whatever you like with—we have no preferences about what you do with the money. You should use it however

you think best."). Consent was obtained before transfers were made. Transfers were sent using M-Pesa, a mobile money transfer system operated by the mobile provider Safaricom which allows people to deposit and withdraw money at numerous locations throughout the country. The majority of respondents in our sample had personal M-Pesa accounts. Respondents who did not have a personal M-Pesa account (usually due to not owning a phone) were offered phones which were delivered to their homes, with the retail price (KES 1600) deducted from their cash transfers. They were guided on how to set up an M-Pesa account, and later revisited to collect the names and phone numbers with which they had registered. This information was then cross-checked with data from Safaricom (which allows checking with name is associated with a given number), and recipients were told that they would receive the (first) cash transfer within 5 weeks after this confirmation. Transfers were made within the appropriate week, conditional on the number and names having been confirmed before the start of that week.[12] Cash transfers were yoked to PM+ delivery to ensure that timing was matched across treatment groups. Respondents in CT&PM+ villages were treated simultaneously with both the CT and PM+ groups. Cash transfers were randomly delivered either as lump-sum transfers of KES 50,000, or in 5 weekly installments of KES 10,000 each week for 5 weeks, to match the delivery of PM+. The timing of lump-sum transfers was randomized within the five-week period of the installments for the same cohort (e.g. if installments for a cohort started in week 29, lump-sum transfers within that same cohort were equally likely to happen in weeks 29, 30, 31, 32, or 33). Compliance was high, with 98.1 percent of participants in the CT and CT&PM+ treatments receiving their transfers according to the M-Pesa receipts sent to Busara by Safaricom.

Again we independently verified these numbers 18 months later by calling recipients. As above, we tried to call all 493 CT&PM+ recipients. We additionally targeted a random subset of 100 households from the CT group.

[12] In case of discrepancies between the respondent names and the account names registered with Safaricom, the respondent was contacted again and the names and numbers were confirmed until they matched.

We reached 400 CT&PM+ households (81 percent) and 89 CT recipients (89 percent). 94 percent of respondents reported no issues receiving their transfers. The remaining 6 percent recalled receiving no transfer or less than the promised amount. However, 82 percent of respondents who claimed they did not receive a transfer did in fact receive it, as per Busara's M-Pesa receipts. Busara's records confirmed that 4 respondents did not receive their transfer, because they either did not consent, had their M-Pesa account suspended, or did not hold an identification document needed to register with the M-Pesa platform. Like for PM+, we use intent-to-treat for these participants.

I.D Data and variables

I.D.1 Survey instruments

A baseline survey was conducted between October 2016 and March 2017, and an endline survey was conducted between August 2018 and May 2019. Online Appendix Table A.1 and Figure A.1 show an overview of the time elapsed, in months, between baseline and intervention start, intervention start/end and endline, and baseline and endline. The mean delay between the end of interventions and endline was 12.55 months, with a median of 13 months. The mean delay between the start of interventions and endline was 13.97 months (median: 14 months).

Both at baseline and endline, we conducted two surveys in each household: a household survey, and an intimate partner violence (IPV) survey. The household survey contained modules on consumption, food security, assets, revenue, profits, labor, education, and psychological well-being. The IPV survey contained questions on the prevalence of IPV, and norms around IPV. The household survey was given to the main respondent, as described above. At endline, the same respondent was re-surveyed. Note that this was the same person who received the intervention(s) in the treatment group. The IPV survey was given to female respondents who were cohabiting with or married to a man; and to the female spouses or cohabiting partners of male respondents, on a different day than the household survey. It was always conducted by a

female FO, who took special care to ensure privacy due to the sensitive nature of the questions.

I.D.2 Variables

The full list of outcome variables is shown in Online Appendix Section I. Following our pre-analysis plan, we grouped our outcomes into four primary categories, and five secondary categories. The primary outcome categories were consumption, asset holdings, revenue, psychological well-being, and intimate partner violence. The secondary outcome categories were food security, profits, labor, and education. For each category, we computed a pre-specified index variable. In the case of monetary variables, this was usually the sum over a number of component variables, such as total household consumption or total asset holdings. For other variables it was a standardized weighted average following Anderson (2008). Monetary variables are winsorized at 1 percent and 99 percent.

Our primary outcomes are the following variables:

1. Monthly per-capita non-durable consumption (USD PPP): This variable measures total monthly expenditure on non-durables per capita, and the value of food consumption from own production, in USD PPP, dividing total household consumption by the number of household members (i.e. considering children as full adult equivalents). Sub-categories are food consumption (from own production and food bought); temptation goods; personal and household items; housing repair and improvement; and education expenditure, medical care, and social activities. The survey contained 134 food expenditure categories (recall horizon one week), 56 common non-food expenditure categories, such as airtime and firewood (recall horizon one month), and 9 low-frequency expenditure categories, such as weddings and school fees (recall horizon 12 months).

2. Total value of assets owned (USD PPP): This variable measures the total value in USD PPP of assets owned by the household. Sub-categories are productive assets, such as wheelbarrows and farming tools; vehicles,

such as bicycles and motorbikes; furniture; household durables, such as cell phones and kerosene stoves; livestock; and financial assets, i.e. savings and debt. The survey covered a total of 30 asset categories, and measures current value as estimated by the respondent (rather than purchase price).

3. Monthly household revenue (USD PPP): This variable measures total household revenue in USD PPP. Sub-categories are revenue from livestock (e.g. milk and meat sales); income from crop sales; enterprise income (e.g. from kiosks); and wage income (e.g. from salaried jobs or casual labor).

4. Subjective well-being index (standard deviations): This variable is a standardized weighted average of several individual scales and questions: the General Health Questionnaire (GHQ-12; Goldberg and Blackwell 1970), a 12-item screener for general psychiatric conditions; the Perceived Stress Scale (PSS; 4 items, Cohen, Kamarck, and Mermelstein 1983); and the Happiness and Life Satisfaction questions from the World Values Survey (WVS). The standardized weighted average is computed following Anderson (2008), i.e. by overweighting variables which are less correlated with others and therefore presumably add more information. The WHODAS Disability Assessment Schedule 2.0, which measures difficulties due to health conditions and is thus a function assessment, was analyzed as an additional, non-primary outcome that did not enter the index. The GHQ-12 has been used extensively to assess psychiatric morbidity in low- and middle-income countries, including Kenya, where it has been validated and used in previous randomized controlled trials, including that by Bryant et al. on PM+ (Abubakar and Fischer 2012; Getanda, Papadopoulos, and Evans 2015; Bryant et al. 2017). The PSS and WVS items have also been previously validated and used in Kenya (Haushofer and Shapiro 2016; Haushofer et al. 2020), and the same is true for the WHODAS (Bryant et al. 2017; Cresswell et al. 2020). The scales had good properties in our sample, with Cronbach's alpha values

of 0.88 for GHQ-12, 0.60 for PSS, and 0.78 for WHODAS.

5. Intimate partner violence index (standard deviations): This variable is a standardized weighted average of two indices, one for physical and one for sexual violence. The physical violence index is a standardized weighted average of 9 variables measuring the frequency over the preceding 6 months of various forms of violence perpetrated by the male intimate partner (usually the husband) against the female respondent, and against any children under 12 in the household. The sexual violence index is a standardized weighted average of variables measuring the frequency of rapes and other forced sexual acts committed by the intimate partner against the respondent over the preceding 6 months.
   In explicit self-reports of IPV, we may worry about over- or under-reporting due to demand effects, e.g. arising from the perceived desirability of a particular response. We therefore supplement the explicit self-reports of IPV with two additional measures. First, an "envelope task" presented participants with 3 envelopes, of which two contained a yellow and one a red piece of paper. Participants were instructed to shuffle the envelopes and, without the surveyor watching, choose one and privately observe the color of the piece of paper it contained. They were then given two yes/no questions: one which they were supposed to answer if their envelope contained yellow paper, and another if their envelope contained red paper. Crucially, because the surveyor did not know the color of the paper, they did not know to which question the participant was responding. There were two rounds of this task. In the first round, the yellow question was: "In the past six months, have you ever visited a hospital or clinic?" The red question was: "In the past 6 months, has your husband ever beaten you, slapped you, or acted violently against you?" In the second round, the red question was: "In the past 6 months, have you purchased any insurance or fertilizer?" The yellow question was: "In the past 6 months, has your husband ever forced you to have sexual intercourse with him even when you did not want to?" In the analysis, we simply record the number of "yes" responses, knowing

that in expectation half of them are in response to the non-sensitive and the other half in response to the sensitive question. Decreases in violence would be reflected as decreases in the share of yes responses.

Second, in the "smiley task", participants were presented with a sheet of paper with a happy face and a sad face on it, and asked to point to the sad face if the husband had beaten, slapped, or acted violently against them; and if the husband ever forced them to have sexual intercourse with him even when they did not want to in the past 6 months. We record whether participants point to the sad face, separately for each of these two questions.

The envelope task was presented first; the smiley task second; and the explicit questions last.

Our secondary outcomes were defined as follows:

1. Food security (standard deviations): This variable is a standardized weighted average of 6 variables measuring food security, such as the frequency over the preceding month of skipping meals, borrowing food from others, and eating protein.

2. Profits (USD PPP): This variable measures total household profits, i.e. revenue minus costs, from livestock, crops, and enterprises. Revenue is usually from sales, and costs include fodder, veterinary care, seeds, fertilizer, hired labor, etc.

3. Labor supply (hours): This variable measures the hours spent per week per capita on income-generating activities, including working in agriculture and tending animals, working in a non-farm or livestock business, and working for others. The per capita figure is obtained by dividing the total number of hours supplied by the household by the number of adults (older than 18) in the household.

4. Education (standard deviations): This variable is a standardized weighted average of the proportion of children enrolled in school; the average days of school missed per child in the preceding 30 days; the average spending

on school expenses per child in the preceding 12 months; and the average time spent studying or in school per child in the preceding 7 days. The variable is defined only for households with school-age children (between 5 and 19 years old).

The questionnaire contained a number of additional variables which were not pre-specified to be primary or secondary. Definitions for these variables are shown in the Online Appendix.

I.D.3 Data quality

Surveys were administered on tablet computers using the *SurveyCTO* survey software. Data integrity was monitored through a series of checks: high-frequency checks, which checked incoming data for completeness on a weekly basis; back-checks, in which all survey respondents were re-surveyed within a week of the original survey to confirm a subset of presumably immutable responses, such as age or number of children; random spot checks in the field, i.e. visits and monitoring of surveys by supervisors; GPS checks, in which Google Earth was used to confirm the existence of a dwelling at the GPS locations recorded in the survey; and M-Pesa confirmation, which consisted of confirming with Safaricom that the name of the respondent matched the name of the M-Pesa account. Further details are given in Online Appendix Section A.1.

II. Econometric specifications

II.A Direct and spillover effects

We use the following ANCOVA framework as our primary specification (McKenzie 2012):

$$\begin{aligned} y_{vi} = \beta_0 + \beta_1 CT_{vi} + \beta_2 PMP_{vi} + \beta_3 CT\&PMP_{vi} + \beta_4 DFS_{vi} \\ + \beta_5 SpillCT_{iv} + \beta_6 SpillPMP + \gamma' \mathbf{X}_{vi} + \delta y_{viB} + \varepsilon_{vi} \end{aligned} \tag{1}$$

Here, y_{vi} is an outcome for household i in village v; CT_{vi}, PMP_{vi}, and $CT\&PMP_{vi}$ are indicator variables for whether household i in village v received a cash transfer, PM+, or both, respectively. $SpillCT_{vi}$ and $SpillPMP_{vi}$ are indicator variables for whether the household is a spillover household in either cash transfer or PM+ villages. DFS_{vi} is an indicator for whether the household received an incentive to register for a mobile money account.[13] $\mathbf{X}_{vi}$ is a vector of stratification variables, including dummies for being female, having M-Pesa access, and being above the median at baseline on the psychological well-being index, the asset index, the IPV index, and village size. y_{viB} is the outcome variable at baseline; it was not collected for all outcomes, in which case it is omitted. Where baseline data was collected but is missing for some observations, we code the missing value as zero and include a separate dummy variable indicating such replacements. Standard errors are clustered at the village level. The omitted category is households living in pure control villages who did not receive an incentive to register for a mobile money account. Note that the DFS treatment selected only households without M-Pesa access at baseline, and therefore including this indicator effectively removes a number of respondents without M-Pesa at baseline from the sample. We therefore overweight respondents without M-Pesa accounts accordingly. The coefficients β_1, β_2, and β_3 identify the treatment effects of the cash transfer, PM+, and combined intervention, respectively, relative to control. To test whether the interventions are complements, we test whether the sum of the treatment effects of the two individual interventions is smaller than that of the combined intervention, $\beta_1 + \beta_2 < \beta_3$. The coefficients β_5 and β_6 test whether cash transfers and PM+, respectively, have spillover effects.

To account for multiple hypothesis testing, we report False Discovery Rate (FDR) adjusted standard errors, correcting across our five primary outcome variables. We do not correct standard errors for multiple hypothesis testing for our secondary outcomes; for the individual variables that make up the primary and secondary outcomes; and for all other variables in the survey.

[13] We do not report this coefficient; it is the subject of a separate study.

II.B Baseline balance and attrition

To test for baseline balance, we estimate our standard treatment effect equation 1 using baseline variables as outcomes (omitting baseline controls on the right-hand side). The baseline variables are constructed as follows: The baseline asset index is a standardized weighted average of responses to a shortened asset module administered at baseline. This module asked respondents if their household owns their home, land, a bed, mattresses, cell phones, cattle, oxen, bulls, goats, sheep, and savings. The baseline psychological well-being index is a standardized weighted average of scores on the GHQ-12, stress, and WHODAS questionnaires. The baseline indices on IPV and justifiability of violence are constructed in the same way as their endline equivalents. Finally, we include variables for gender, age, and M-Pesa access at baseline.

To test for differential attrition, we estimate the following model:

$$y_{vi} = \beta_0 + \beta_1 CT_{vi} + \beta_2 PMP_{vi} + \beta_3 CT\&PMP_{vi} + \gamma X_{vi} + \delta y_{viB} + \varepsilon_{vi}$$

Here, y_{vi} is an indicator for participant i in village v completing both surveys. The coefficients on CT_{vi}, PMP_{vi}, and $CT\&PMP_{vi}$ therefore measure differential attrition across these treatment groups relative to pure control. Standard errors are clustered at the village level. To assess within-village differences in attrition, we include village-level fixed effects and restrict the sample to cash-only and PM+-only villages. Standard errors are not clustered.

II.C Heterogeneous treatment effects

We test whether the effects of our treatments vary along the following prespecified dimensions: baseline asset holdings; baseline intimate partner violence index; gender of recipient; and baseline psychological distress. For recipient gender, the dimension of heterogeneity is a dummy variable for the main survey respondent being female. For the other dimensions, we create an indicator variable based on a median split. We then estimate estimate the

differential impact by including a dummy for the interactant, and interaction terms with all treatment arms, in the main specification (see Online Appendix Section D for details). Baseline psychological distress is measured as an index composed of the WHO Disability Assessment Schedule (WHODAS) 2.0 and the GHQ-12, which are the measures used by the implementing NGO to diagnose psychological distress. In addition to the pre-specified median split cutoff on this measure, we also conduct an exploratory analysis in which we use the cutoff according to which the implementing NGO diagnoses individuals as having high distress. According to this criterion, individuals are classified as having high distress if their WHODAS score is greater than 2, which is the case for 40 percent of our participants at baseline; and having a GHQ-12 score above 16, which is the case for 30 percent of baseline respondents. Together, 33 percent of participants would have been classified as distressed at baseline according to the NGO's criteria.

II.D Demand effects

Our questionnaires rely on self-reports to measure treatment effects. This raises the concern that participants may alter their responses to "please" the experimenters, potentially leading to biased treatment effect estimates. To investigate whether such "experimenter demand effects" play a role in our study, we use the method proposed by De Quidt, Haushofer, and Roth (2018). This method bounds demand effects by deliberately inducing them through "demand treatments". In these treatments, participants are explicitly told what the experimenters expect of them. We delivered such demand treatments to all respondents in the pure control group at endline, for three questions pertaining to physical violence, sexual violence, and depression. Each pure control respondent was allocated to either a "positive demand" or "negative demand" treatment. Respondents in the "positive demand" treatment are asked the following questions:

1. "We will now ask you questions about how your partner has acted towards you in the last 6 months. We hypothesize that people who participated

in this study and received the same treatment as you will give higher responses to these questions than others. How many times per month did your husband beat you, slap you or act violently against you?"

2. As above, but with the final question being: "How many times per month did your husband physically force you to have sexual intercourse with him even when you did not want to?"

3. "I will read out a list of some of the ways you may feel or behave. Please indicate how often you have felt this way during the past week, using the following scale: Rarely or none of the time (<1 day); Some or little of the time (1-2 days); Occasionally or a moderate amount of time (3-4 days); All of the time (5-7 days). We hypothesize that people who participated in this study and received the same treatment as you will give higher responses to these questions than others. I felt depressed."

Respondents in the "negative demand" treatment were asked the same questions, except that the word "higher" was replaced with "lower". The main idea behind the method by De Quidt, Haushofer, and Roth (2018) is experimenter demand is not a major issue if the "positive demand" and "negative demand" responses are similar. We test this by averaging the responses to the positive and negative demand questions and testing the difference in means using a t-test.

III. Results

III.A Baseline balance and attrition

We report estimates of baseline balance in Online Appendix Table A.2, and of attrition in Online Appendix Table A.3. We find no significant differences between these outcomes at baseline between any of our treatment groups and the control group. For the spillover groups, two coefficients are significant at the 10 percent level (psychological well-being was somewhat higher in the PM+ spillover group compared to control, and the justifiability of violence

somewhat lower in the cash spillover group compared to control). However, these differences are weak and the number of tests conducted relatively large, and we therefore do not regard this as evidence of imbalance. Our randomization thus appears to have been successful in creating comparable treatment groups.

Online Appendix Table A.3 shows that we had 5 percent attrition between baseline and endline in pure control and PM+ villages, and 4 percent in cash and "Cash & PM" villages. The average completion rate of the endline survey for participants who completed baseline was 96 percent. These levels of attrition are low and not statistically different from each other. Comparing treatment to spillover households within cash-only and PM+-only villages (columns (2) and (3)) also shows no differential attrition within treatment villages. Thus, we had low attrition overall, and it was not differential.

III.B Treatment effects on primary and secondary outcomes

We next turn to treatment effects on our primary outcomes of interest, shown in Table 1 and Figure 3. Detailed tables are shown in Online Appendix Section B.1. Column (1) of Table 1 shows the pure control group mean and standard deviation; for standardized variables, the mean is zero and the standard deviation 1 by definition. The cash transfer treatment has positive and statistically significant effects on 4 of our 5 primary outcomes, shown in column (2): consumption, asset holdings, revenue, and subjective well-being. The effect on consumption is a USD 10.51 PPP increase relative to a control group mean of 52.49, corresponding to a 20 percent increase. Similarly, the effect on assets is an increase of USD 262.06 PPP relative to a control group mean of 553.16, or a 47 percent increase. Revenue increases by USD 35.18 PPP per month, a 26 percent increase relative to the control group mean of USD 135.48 PPP. The subjective well-being index increases by 0.23 standard deviations (SD). All of these effects are significant at the 1 percent level using both naïve and FDR-adjusted p-values. We observe no significant effect on the intimate partner

violence index, which decreases by 0.05 SD.

The PM+ treatment, shown in column (3), has no significant effects on any of our primary outcomes except for intimate partner violence, where we observe an increase (i.e. a worsening of violence) of 0.24 SD. However, two caveats apply to this effect. First, it does not survive FDR correction. Second, as shown in Online Appendix Table B.5, when we use the envelope and smiley tasks to measure prevalence of IPV, we observe no significant increases in IPV. Thus, the impacts of the PM+ program on IPV are inconclusive. Nevertheless, future work should take seriously the possibility that PM+ might increase IPV, and attempt to provide more robust evidence and study possible mechanisms.

The point estimates on consumption, assets, and revenue are positive, with a 5 percent increase in both consumption and asset holdings, and a 9 precent increase in household revenue. However, these effects are not statistically different from zero using both naïve and FDR-adjusted p-values. Using the 95 percent confidence intervals, we can rule out treatment effects larger than 18 percent on consumption, 21 percent on asset holdings, 24 percent on revenue, and 0.11 SD percent on psychological well-being. Column (5) reports the p-values testing the difference between the effects of the cash treatment and the PM+ treatment; all tests are statistically significant, suggesting that the cash treatment had significantly larger positive effects than the PM+ treatment.

Column (4) reports the treatment effect of the combined cash and PM+ treatment. These treatment effects are both numerically and statistically very similar to those of the cash-only treatment; for example, asset holdings increases by USD 227.61 PPP in the cash & PM+ group, a 41 percent increase relative to the control group mean, and similar to the USD 262.06 PPP treatment effect in the cash-only treatment. Similarly, revenue increases by USD 23.00 PPP, not too dissimilar to the USD 35.18 PPP increase in the cash-only treatment. The coefficient on the subjective well-being index is 0.27 SD, compared to 0.23 SD in the cash treatment. Only consumption shows a smaller effect, with a USD 3.62 PPP treatment effect in the combined treatment, compared to a USD 10.51 PPP treatment effect in the cash-only treatment. In the combined treatment, only the impacts on assets, revenue, and subjective

well-being are statistically significant (using both naïve and FDR-adjusted p-values), and the impact on revenue is only significant at the 10 percent level. Column (6) shows the difference between the cash treatment and the combined treatment, revealing that the impact on consumption is smaller in the combined than in the cash-only treatment (significant at the 10 percent level); the other effects are statistically similar.

Thus, the combined treatment has similar effects to the cash-only treatment, even though they are statistically somewhat less robust, and the consumption effect may in fact be smaller. All impacts in the combined treatment are numerically larger than those of the PM+ treatment, although as shown in column (7), only asset holdings and subjective well-being show statistically significant differences across the two treatment arms.

Finally, column (8) shows p-values comparing the sum of the treatment effects of the cash-only and PM+-only treatment arms to the effects of the combined treatment. With the exception of the consumption outcome, none of the p-values are statistically significant at conventional levels, which is not surprising given the small effects of the PM+ treatment and the similar effects of the cash-only and combined treatments. The summed effect on consumption is larger than the effect in the combined treatment arm, significant at the 10 percent level.

Table 2 reports effects on our secondary variables of interest. Mirroring the positive treatment effect of the cash transfer on consumption, we observe a 0.13 SD improvement in the food security index in this treatment arm, significant at the 5 percent level. We observe a 27 percent decrease in profits, but this effect is not statistically distinguishable from zero. Together with the positive effects on revenue, it suggests that cash transfer recipients grew their business, rather than increasing their profits. We see a small and insignificant reduction in labor supply, and a negligible effect on the education index. None of these effects are statistically significant, even with naïve p-values.

The PM+ treatment also has small and insignificant treatment effects on the secondary outcome variables, with numerically small reductions in food security, profits and working hours, and a small increase in the education

index. Again no result is statistically significant.

The combined cash and PM+ treatment also shows a significant increase in the food security index (0.12 SD). Apart from that, no coefficients are statistically significant, although there is a numerically relevant decrease in profits (−37 percent). Again this result corroborates the pattern that cash transfer recipients grew their businesses rather than increasing their profits. The other treatment effects are numerically small.

In line with the generally small impacts amongst these secondary outcome variables, the pairwise comparisons shown in columns (5)–(8) are not statistically significant for the most part, with the sole exception of the food security index, which larger in the two treatment arms receiving cash than in the pure PM+ group.

III.C Treatment effects on psychological well-being

Table 3 reports the impacts on the individual psychological well-being variables, all coded in standard deviations. The 0.23 SD effect of the cash transfer treatment on the subjective well-being index is driven by a 0.21 SD impact on life satisfaction; a 0.19 SD increase in happiness; and 0.16 SD decreases in the perceived stress and GHQ-12 scales. All effects are statistically significant. In contrast, the PM+-only treatment has no significant treatment effects on any of these variables, and all individual coefficients are smaller than 0.05 SD. The combined treatment has a similar effect on the psychological well-being index (0.27 SD) as the cash-only treatment (0.23 SD), and most individual impacts are also numerically similar. However, we observe a larger impact on life satisfaction in the combined treatment (0.39 SD) relative to the cash treatment (0.21 SD), with the difference significant at the 10 percent level. Thus, for this specific dimension, combining cash transfers with PM+ may have increasing returns, but the result needs replication. The function assessment, WHODAS, did not show significant treatment effects in any treatment arm.

Detailed treatment effects on other outcome groups are shown in Online Appendix Section B.1.

III.D Interpreting null results: A Bayesian Approach

We briefly discuss what inferences can be drawn about the effectiveness of the PM+ program based on the largely null results we report above. For illustrative purposes, we focus on the effect of the PM+ program on the psychological well-being index. We first note that the standard error of the treatment effect is 0.06 SD, which means that we had 80 percent power to detect effect sizes of $2.8 \times 0.06 \approx 0.17$ SD at a significance level of 5 percent.[14] This detectable effect size is relatively small compared to treatment effects of psychotherapy reported throughout the literature. For instance, Cuijpers et al. (2010) report an average treatment effect of 0.67 SD in a meta-analysis of 117 randomized experiments comparing psychotherapy interventions to control conditions, which is still 0.42 SD after statistically adjusting for publication bias. For cognitive-behavioral therapy specifically, on which PM+ is based, Cuijpers et al. (2013) report a meta-analytic effect size estimate of 0.71 SD (no adjustment for publication bias was undertaken). In the previous PM+ trial in Kenya, Bryant et al. (2017) report a 0.57 standard deviation (SD) improvement in the GHQ-12. Even allowing for the fact that PM+ is a low-dose, highly manualized version of CBT delivered by laypeople, the fact that we were powered to detect much smaller effects makes power an unlikely explanation of our findings.

A natural question to ask is then what the probability is that the true effect of the PM+ program is smaller than 0.17 SD, our detectable effect size. This probability is given by the *negative predictive value* (NPV): the probability that there is no true effect of a given magnitude, given that no effect was detected with a given level of power and false positive probability. In addition to power and the false positive rate, the negative predictive value is influenced by priors, i.e. the probability we would have assigned to PM+ having an effect of at least 0.17 SD before the study. For power $1 - \beta$ (and thus false

[14] This insight uses the fact that to declare significance at 5 percent, the absolute value of the test statistic must be larger than 1.96, and for this to be true with 80 percent probability, 80 percent of the sampling distribution from which test statistics are drawn must be to the right of 1.96. Because the 80 percent of the standard normal distribution is given by a z-score of 0.84, this is the case when the sampling distribution is centered $1.96 + 0.84 = 2.8$ standard errors away from zero.

negative rate β), false positive rate α, and prior probability π, the NPV is given by Bayes' theorem as the ratio of the expected rate of false negatives, $(1-\alpha)(1-\pi)$, to the expected rate of negative results overall, $(1-\alpha)(1-\pi)+\beta\pi$:

$$NPV = \frac{(1-\alpha)(1-\pi)}{(1-\alpha)(1-\pi)+\beta\pi}$$

In our case, $\alpha = 0.05$, and we calculated above that for an effect size of 0.17 SD, $1-\beta = 0.8$. If we were completely agnostic as to whether PM+ is effective prior to the study (flat priors, $\pi = 0.5$), the post-study probability of PM+ having a true effect smaller than 0.17 is 0.83. With a prior of 90 percent in favor of PM+, the post-study probability against PM+ is 0.35. Thus, depending on the strength of our priors, the post-study probability of PM+ being ineffective is only moderately increased by our results. However, larger effects can be ruled out with greater confidence: even with a 90 percent prior in favor of its effectiveness, the post-study probability that PM+ produces effects smaller than 0.40 SD is 90 percent.

III.E Cost effectiveness

The total amount paid to the NGO for the delivery of PM+ to the 1018 recipients (525 recipients in the PM+-only group, and 493 recipients in the combined group) was USD 1,210,107 (nominal), which includes both direct program cost and overhead. This corresponds to a program cost of USD 1,189 (nominal) per recipient. This cost is larger than the nominal amount of the cash transfers, USD 485, by a factor of 2.45. We conservatively assume that the costs of remitting the cash transfers, in M-Pesa and staff costs, amounted to 10 percent of the transfer value, for a cost of USD 534 per transfer. Using this cost, the PM+ program is more expensive by a factor of 2.27. The PM+ facilitators were paid KES 700 per session of PM+ administered, for a total of KES 3500 (USD 37 nominal, USD 75 PPP) across the five sessions of the program. At this marginal cost for PM+, the nominal amount of the cash transfer exceeds the marginal cost of PM+ by a factor of 14.43. The point

estimates of the cash transfer are larger than those of PM+ by factors of 4.17 (consumption), 10.04 (assets), and 3.04 (revenue). It is possible that the PM+ program would be substantially cheaper in the future, when infrastructure is already in place and CHWs are already trained; under such conditions, PM+ may be more cost-effective than cash transfers for these outcomes.

Online Appendix Figure C.1 illustrates the relative cost effectiveness of PM+ and cash transfers at different scenarios for the cost of PM+. For the purposes of this analysis, we take seriously the statistically non-significant, but positive impacts of PM+ on our monetary outcomes, i.e. consumption, assets, and revenue. (Note that for the other primary outcomes, the PM+ impacts have negative signs, while those of cash transfers are positive, making PM+ less cost-effective than cash transfers in all cost scenarios.) The horizontal axis represents different possible costs of delivering PM+ to a single recipient, ranging from marginal cost (USD 37 nominal) to total cost in this study (USD 1,189 nominal). On the vertical axis, we plot the ratio of the treatment effect of PM+ on monetary outcomes, per dollar spent on PM+, relative to the treatment effect of cash transfers on the same outcomes, per dollar spent on cash transfers (i.e. $\frac{\hat{\beta}_2/c_{PMP}}{\hat{\beta}_1/534}$, where c_{PMP} is the varying cost of PM+, and $\hat{\beta}_1$ and $\hat{\beta}_2$ are the cash and PM+ treatment effect estimates from our main estimating equation, respectively). The dotted vertical lines represent the PM+ cost below which PM+ is more cost-effective in changing the outcome variable in question than cash transfers. We find PM+ may have the potential to be more cost-effective than cash transfers in improving revenue if the PM+ cost can be reduced to USD 175 per person; USD 128 for consumption; and USD 53 for assets.

III.F Heterogeneous treatment effects

We report heterogeneous treatment effects of the cash transfer treatment, PM+ treatment, and the combined treatment on our primary outcomes in Online Appendix Tables D.1, D.2, and D.3, respectively. In Online Appendix Table D.1, we find that cash transfers generate a somewhat smaller effect on

expenditure in female recipient households compared to male recipient households. Note, however, that recipient gender was not randomized, and thus these households may differ in unobserved ways. We also find that for individuals with high psychological distress at baseline, cash transfers have a substantially smaller impact on revenue than in other households. This result is in line with the view that psychological distress may make it difficult to put cash transfers to productive use. We find suggestive evidence that transfers to asset-rich households increase IPV more strongly than transfers to other households, and that asset holdings increase less strongly in households with high baseline IPV. However, these effects are somewhat imprecise and we therefore do not interpret them strongly.

We find no evidence of interaction effects of the PM+ treatment with any of our pre-specified dimensions of heterogeneity (Online Appendix Table D.2); none of the interaction terms reach statistical significance. Most prominently, PM+ effects on psychological well-being are not larger in participants who experienced high psychological distress at baseline. Recall that the inclusion criteria for the study did not include having psychological distress at baseline. Thus, the small treatment effects of the PM+ treatment on psychological well-being outcomes (not to mention economic outcomes) might result from PM+ not being very effective for people who are not psychologically distressed. However, as shown in columns (4) and (5), the interaction terms are numerically not large and not statistically significant. This is true both for a median split on baseline psychological distress (column 5), and when the cutoff for baseline distress is that typically used by the implementing NGO (column 4). Thus, PM+ has limited effectiveness even for psychologically distressed recipients in our setting.

In a further exploratory analysis, we next ask if perhaps PM+ affected only the *most* severely distressed participants, even amongst those classified as distressed. To this end, we consider the 33 percent of participants who were classified as distressed at baseline (based on the index constructed from GHQ-12 and WHODAS scores), and further split this subsample into "distressed" and "severely distressed" participants. We do this for every possible

cutoff for "severe distress", beginning with the full 33 percent, and then gradually increasing the cutoff until the heterogeneity analysis is based only on the single most severely distressed participant. We estimate interaction coefficients and confidence intervals at each point. The results are presented in Online Appendix Figure D.1. We do not find any evidence for heterogeneous PM+ treatment effects on the "severely distressed" sub-subgroup, for any of our outcomes, at any chosen split point.

Because PM+ was effective in reducing psychological distress among women who had experienced gender-based violence in a previous study in Kenya (Bryant et al. 2017), we might expect that women, or women who reported high levels of IPV at baseline, would show larger impacts. However, in columns (2) and (3) of Table D.2, we find no evidence for larger effects of PM+ on the psychological well-being index of women, and of women who experienced high levels of IPV (in fact, both coefficients are negative, although not statistically significant).

Finally, in Table D.3, we report heterogeneous treatment effects of the combined cash and PM+ treatment. We find that in households with high baseline IPV, this treatment had a larger (yet imprecisely estimated) positive effect on revenue than in other households; possibly because in these households, husbands were more likely to appropriate transfer money (and transfers to husbands are often more likely to be invested in business; cf. De Mel, McKenzie, and Woodruff 2012). At the same time, asset holdings increase more strongly in female recipient households compared to male recipient households, although this effect is similarly imprecisely estimated.

III.G CHW fixed effects

To understand whether some CHWs were more effective than others in delivering PM+, we proceed as follows. First, we estimate separate treatment effects on the psychological well-being index by CHW. Specifically, we restrict the sample to the PM+-only and the pure control group, and regress the z-scored psychological well-being index on indicator variables for each of the CHWs who

delivered PM+ (recall that the assignment of CHWs to clients was random). A histogram of the resulting coefficients, each of which can be considered the impact of receiving PM+ treatment from a specific CHW, is shown in Figure E.1. We find a fair degree of variability in the treatment effect as a function of CHW, suggesting that some CHWs were more effective than others in administering PM+. Second, to ensure that this pattern reflects more than natural variability in outcomes, we ask whether CHW performance predicts outcomes out-of-sample by testing whether the outcomes of each PM+ participant are predicted by the outcomes of the other participants who were served by the same CHW. Specifically, we regress the endline psychological well-being index of each PM+ participant on the average endline psychological well-being index of the other PM+ participants who were served by the same CHW, clustering standard errors at the CHW level. Note that this approach corresponds to a leave-one-out analysis and is therefore an out-of-sample test of CHW quality. We find a coefficient of 0.34 (standard deviations), which is highly significantly different from zero. Thus, some CHWs were indeed more effective than others.

III.H Treatment effects over time

One possible explanation for the small impact of the PM+ treatment in this study, despite its success in earlier studies (Bryant et al. 2017), is that the program only has short-lived effects that subside over time. Our endline was on average 13 months after the end of the interventions, whereas the previous demonstrations of effectiveness happened 3 months after the intervention. We did not have a short-run endline; however, we randomized the survey timing and therefore have variation in the time elapsed between the end of interventions and the endline. In the following exploratory analysis, we therefore investigate how treatment effects change over time. In particular, we are interested to see if impacts are strongest right after treatment administration. To answer this question, we estimate our main treatment equation 1 over gradually expanding subsets of data. Specifically, we begin by restricting the sample to the pure control group and those participants in the other treatment arms

that received their most recent cash transfer or PM+ session 200 or fewer days ago. We then gradually widen this time window by one day and include people whose most recent treatment administration was 201 days ago, then 202 days, and so forth, until we have included the entire sample. The pure control group is always included in its entirety. For participants in the CT treatment, we use the duration in days between the endline survey date and the most recent M-Pesa cash transfer. Similarly, for participants in the PM+ treatment, we use the interval between endline and the most recent PM+ session. For the CT&PM+ arm, we use the most recent date of either a transfer or PM+ session. To control for potential seasonality effects, we include time fixed effects for each individual survey month. Intuitively, we are thus comparing participants in the pure control group to a subset of treated people who were surveyed in the same month, and whose last treatment was at least a given number of days ago.

Online Appendix Figure F.1 shows these moving window response functions for all primary and secondary outcomes. In each of these nine panels, the horizontal axis corresponds to the time elapsed between endline and the most recent cash transfer or PM+ session. The vertical axis shows treatment effects estimated for the sample defined by each individual point on the horizontal axis. PM+ treatment effects are shown in red, cash effects in green, effects from the combined treatment in blue. Note that the rightmost edge of each of these graphs (about 700 or fewer days) includes the entire sample, and hence corresponds to the treatment effects of our main tables.

We find that the effects of cash transfers on consumption and assets begin somewhat larger and then go down over time, although they are still large and positive when considering the entire sample (and hence all time horizons). Psychological well-being is elevated at all time horizons for cash transfer recipients. Importantly, however, well-being is not elevated in the PM+-only treatment, even at short time horizons. Thus, it is unlikely that PM+ had early effects which then subsided over time. We note, however, that even our shortest time horizon—200 days or about 7 months—is longer than that considered in the previous study demonstrating the effectiveness of PM+ in

Kenya (Bryant et al. 2017). It therefore remains possible that the program would show larger effects at even shorter time horizons, even though the decline between 3 and 7 months would have to be rather steep.

III.I Demand effects

Experimenter demand effects are a potential concern when using self-report measures for outcomes of interest. For instance, we might worry that cash transfer recipients want to express their gratefulness for receiving cash transfers by giving more positive responses to questions about subjective well-being than they otherwise would, leading to an observed treatment effect that in actuality reflects a demand effect. However, such effects are unlikely to explain our findings, for three reasons. First, existing evidence on experimenter demand effects suggests that they are not large in magnitude, both in general terms and in relation to the effects we report here (De Quidt, Haushofer, and Roth 2018). Second, we observe treatment effects on some outcome variables (e.g. cash transfers on psychological well-being), but not on others which should be at least equally affected by demand effects (e.g. PM+ on psychological well-being). Finally, we test directly for the importance of demand effects in our sample, using the approach proposed by De Quidt, Haushofer, and Roth (2018) and described in Section II.D. Results are reported in Online Appendix Table G.1. We find no evidence of demand effects; the mean responses given in the "positive demand" and "negative demand" conditions are almost exactly identical: 4.65 vs. 4.62 for the question about beatings; 2.79 vs. 2.76 for the question about rape; and 3.43 vs. 3.42 for the question about depression. The p-values comparing them are not statistically significant. Thus, demand effects appear to play no important role in our setting.

III.J Spillover effects

Online Appendix Tables B.9, B.10, and B.11 show spillover effects for our primary, secondary, and psychological well-being outcomes, respectively. In each table, column (2) shows the spillover effect of cash transfers on non-recipients,

and column (3) that of the PM+ treatment. Column (4) shows the p-values comparing the two spillover effects. We find little evidence of spillovers of cash transfers: consumption is 7 percent higher, and asset holdings 2 percent higher, in non-recipient households in cash transfer villages compared to households in control villages, but the confidence intervals include zero.[15] Revenue and the subjective well-being index are hardly affected at all by spillovers. The intimate partner violence index shows a 0.17 SD increase, suggesting more violence in non-recipient households in cash transfer villages. However, this effect is only significant at the 10 percent level, and does not survive multiple inference correction. We note that this effect goes in the opposite direction to that reported in a previous similar program (Haushofer and Shapiro 2016).

The PM+ treatment also has positive but small spillover effects on consumption and asset holdings, with 3 percent and 6 percent increases, respectively, both not statistically significant. Interestingly, revenue in non-recipient households in PM+ villages is 17 percent higher than in control households; however, this effect does not survive multiple inference correction. We find little evidence of spillovers on psychological well-being. In line with the increase in IPV observed in PM+ households in treatment villages, non-recipient households in these villages also experience an increase in IPV, by 0.16 SD, significant at the 10 percent level. However, this effect, too, does not survive FDR correction. Overall, the spillover effects of our two treatments are thus economically small and statistically weak.

We find no strong evidence for spillover effects on food security, household profits, or working hours. The education index shows a -0.15 SD spillover effect, significant at the 10 percent level.

Because the spillover effects are generally small, the within-village treatment effects (i.e. comparing treatment to spillover households) are similar to the across-village treatment effects. For this reason we do not show detailed tables for this comparison, even though it was pre-specified. We had also pre-specified several analyses to study the mechanisms for possible spillover

[15]Note that, using the approach to compute MDEs described in Section III.D, we were powered to detect relatively small spillover effects.

effects, and we do not show them for the same reason. The results for both sets of analysis are available upon request.

III.K Effects of transfer frequency

We next compare the relative impact of cash transfers when they are made as a lump-sum compared to when they are made in five weekly installments. Details of the econometric approach are shown in Online Appendix Section H. Online Appendix Table H.1 shows the corresponding treatment effects on primary outcomes. Columns (2) and (3) show the treatment effects, relative to control, of weekly and lump-sum transfers, respectively, and column (4) shows the p-value for the difference of these coefficients. We find that weekly transfers have a substantially larger impact on both consumption and revenue than lump-sum transfers: weekly transfers increase consumption by USD 14.98 PPP, significant at the 1 percent level; while lump-sum transfers increase consumption by USD 7.01 PPP (not significant), and the two effects are significantly different from each other. Similarly, weekly transfers increase monthly household revenue by USD 51.71 PPP, while the increase is only USD 19.30 for lump-sum transfers; again the first effect is significant and the second is not, and the difference between them is significant. We find little evidence for a differential impact of weekly vs. lump-sum transfers for asset holdings, subjective well-being, and IPV. We also find no significant differences between weekly and lump-sum transfers on our secondary outcomes (not shown). Thus, weekly transfers have larger positive effects on two of our primary outcome variables, suggesting that in this setting and for these outcomes, they are weakly superior to lump-sum transfers.

IV. Conclusion

In this paper, we report the results of a randomized controlled trial that compared the effects of USD 1076 PPP unconditional cash transfers, a five-week psychotherapy program, and their combination on psychological and economic

well-being among a sample of low-income Kenyans. One year after the intervention, we find improvements in both economic and psychological well-being among cash transfer recipients. In contrast, the psychotherapy program has no measurable positive effects on these outcomes at this time horizon. The results are similar when considering only individuals who had low mental health at baseline, suggesting that these results are not a matter of the psychotherapy program only working for psychologically distressed people. In line with these findings, the effects of the combination of both the unconditional cash transfer and the psychotherapy program are very similar to those of the cash transfer by itself. We observe little evidence of spillover effects of either program, although we find suggestive evidence of spillovers of the cash transfer on intimate partner violence.

Together, these results suggest that in this setting, cash transfers are more effective than this particular psychotherapy program in improving both economic and psychological well-being at a time horizon of one year. This result is important because the cost of the psychotherapy program significantly exceed that of the cash transfer. Thus, *ex post*, a donor might have preferred to invest in cash transfers rather than the psychotherapy program in this study.

It is important to note that the psychotherapy program we used was a state-of-the-art intervention for low-income contexts, developed by the WHO and previously successfully tested in Kenya. In addition, the implementation of the program in our study was done by the same NGO as in this previous study in Kenya. Thus, implementation differences are unlikely to explain the lack of impact. Rather, it appears that this psychotherapy program, even though it is general in nature, may be most effective when it is deployed to address a specific problem such as IPV. In addition, it might be made more effective by increasing its intensity.

References

Abubakar, A., and R. Fischer. 2012. "The Factor Structure of the 12-Item General Health Questionnaire in a Literate Kenyan Population." *Stress and Health: Journal of the International Society for the Investigation of Stress* 28 (3): 248–254.

Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 103 (484): 1481–1495.

Baird, Sarah, Jacobus De Hoop, and Berk Özler. 2013. "Income Shocks and Adolescent Mental Health." *Journal of Human Resources* 48 (2): 370–403.

Bandiera, Oriana, Robin Burgess, Narayan Das, Selim Gulesci, Imran Rasul, and Munshi Sulaiman. 2017. "Labor Markets and Poverty in Village Economies." *The Quarterly Journal of Economics* 132 (2): 811–870.

Banerjee, Abhijit, Esther Duflo, Nathanael Goldberg, Dean Karlan, Robert Osei, William Parienté, Jeremy Shapiro, Bram Thuysbaert, and Christopher Udry. 2015. "A Multifaceted Program Causes Lasting Progress for the Very Poor: Evidence from Six Countries." *Science* 348 (6236): 1260799.

Baranov, Victoria, Sonia Bhalotra, Pietro Biroli, and Joanna Maselko. 2020. "Maternal Depression, Women's Empowerment, and Parental Investment: Evidence from a Randomized Control Trial." *American Economic Review* 110 (3): 824–859.

Bedoya, Guadalupe, Aidan Coville, Johannes Haushofer, Mohammad Isaqzadeh, and Jeremy Shapiro. 2019. "No Household Left Behind: Afghanistan Targeting the Ultra Poor Impact Evaluation." *NBER Working Paper 25981.*

Bernard, Tanguy, Stefan Dercon, Kate Orkin, and Alemayehu Seyoum Taffesse. 2014. "The Future in Mind: Aspirations and Forward-Looking Behaviour in Rural Ethiopia." *BREAD Working Paper* 429:1–42.

Biasi, Barbara, Michael S. Dahl, and Petra Moser. 2019. "Career Effects of Mental Health." SSRN Working Paper 2544251.

Blattman, Christopher, Nathan Fiala, and Sebastian Martinez. 2020. "The Long Term Impacts of Grants on Poverty: 9-Year Evidence from Uganda's Youth Opportunities Program." *American Economic Review: Insights* 2 (3): 287–304.

Blattman, Christopher, Eric P Green, Julian Jamison, M Christian Lehmann, and Jeannie Annan. 2016. "The Returns to Microenterprise Support among the Ultrapoor: A Field Experiment in Postwar Uganda." *American Economic Journal: Applied Economics* 8 (2): 35–64.

Blattman, Christopher, Julian C Jamison, and Margaret Sheridan. 2017. "Reducing Crime and Violence: Experimental Evidence from Cognitive Behavioral Therapy in Liberia." *American Economic Review* 107 (4): 1165–1206.

Bryant, Richard A, Alison Schafer, Katie S Dawson, Dorothy Anjuri, Caroline Mulili, Lincoln Ndogoni, Phiona Koyiet, Marit Sijbrandij, Jeannette Ulate, Melissa Harper Shehadeh, et al. 2017. "Effectiveness of a Brief Behavioural Intervention on Psychological Distress among Women with a History of Gender-Based Violence in Urban Kenya: A Randomised Clinical Trial." *PLoS Medicine* 14 (8): e1002371.

Cohen, Sheldon, Tom Kamarck, and Robin Mermelstein. 1983. "A Global Measure of Perceived Stress." *Journal of Health and Social Behavior* 24 (4): 385–396.

Cresswell, Jenny A., Kelli D. Barbour, Doris Chou, Affette McCaw-Binns, Veronique Filippi, Jose Guilherme Cecatti, Maria Barreix, Max Petzold, Nenad Kostanjsek, Sara Cottler-Casanova, and Lale Say. 2020. "Measurement of Maternal Functioning during Pregnancy and Postpartum: Findings from the Cross-Sectional WHO Pilot Study in Jamaica, Kenya, and Malawi." *BMC Pregnancy and Childbirth* 20 (1): 518.

Cuijpers, Pim, Matthias Berking, Gerhard Andersson, Leanne Quigley, Annet Kleiboer, and Keith S. Dobson. 2013. "A Meta-Analysis of Cognitive-Behavioural Therapy for Adult Depression, Alone and in Comparison with Other Treatments:." *The Canadian Journal of Psychiatry* 58 (7): 376–385.

Cuijpers, Pim, Filip Smit, Ernst Bohlmeijer, Steven D. Hollon, and Gerhard Andersson. 2010. "Efficacy of Cognitive-Behavioural Therapy and Other Psychological Treatments for Adult Depression: Meta-Analytic Study of Publication Bias." *The British Journal of Psychiatry: The Journal of Mental Science* 196 (3): 173–178.

De Mel, Suresh, David McKenzie, and Christopher Woodruff. 2012. "One-Time Transfers of Cash or Capital Have Long-Lasting Effects on Microenterprises in Sri Lanka." *Science* 335 (6071): 962–966.

De Quidt, Jonathan, Johannes Haushofer, and Christopher Roth. 2018. "Measuring and Bounding Experimenter Demand." *American Economic Review* 108 (11): 3266–3302.

Getanda, Elijah Mironga, Chris Papadopoulos, and Hala Evans. 2015. "The Mental Health, Quality of Life and Life Satisfaction of Internally Displaced Persons Living in Nakuru County, Kenya." *BMC Public Health* 15:755.

Ghosal, Sayantan, Smarajit Jana, Anandi Mani, Sandip Mitra, and Sanchari Roy. 2016. "Sex Workers, Stigma and Self-Belief: Evidence from a Psychological Training Programme." *CAGE Working Paper*, no. 302:1–48.

Goldberg, David P, and Barry Blackwell. 1970. "Psychiatric Illness in General Practice: A Detailed Study Using a New Method of Case Identification." *Br Med J* 2 (5707): 439–443.

Hammen, Constance. 2005. "Stress and Depression." *Annu. Rev. Clin. Psychol.* 1:293–319.

Haushofer, Johannes, M. Chemin, Chaning Jang, and Justin Abraham. 2020. "Economic and Psychological Effects of Health Insurance and Cash Transfers: Evidence from a Randomized Experiment in Kenya." *Journal of Development Economics* 144:102416.

Haushofer, Johannes, and Ernst Fehr. 2014. "On the Psychology of Poverty." *Science* 344 (6186): 862–867.

Haushofer, Johannes, Anett John, and Kate Orkin. 2019. "Can Simple Psychological Interventions Increase Preventive Health Investment?" *NBER Working Paper 25731*.

Haushofer, Johannes, Michala Riis-Vestergaard, and Jeremy Shapiro. 2019. "The Comparative Impact of Cash Transfers and Mental Health Interventions." *AEA RCT Registry*, February.

Haushofer, Johannes, and Jeremy Shapiro. 2016. "The Short-Term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya." *The Quarterly Journal of Economics* 131 (4): 1973–2042.

Lund, Crick, Mary De Silva, Sophie Plagerson, Sara Cooper, Dan Chisholm, Jishnu Das, Martin Knapp, and Vikram Patel. 2011. "Poverty and Mental Disorders: Breaking the Cycle in Low-Income and Middle-Income Countries." *The Lancet* 378 (9801): 1502–1514.

Lund, Crick, Marguerite Schneider, Emily C. Garman, Thandi Davies, Memory Munodawafa, Simone Honikman, Arvin Bhana, Judith Bass, Paul Bolton, Michael Dewey, John Joska, Ashraf Kagee, Landon Myer, and Pet. 2020. "Task-Sharing of Psychological Treatment for Antenatal Depression in Khayelitsha, South Africa: Effects on Antenatal and Postnatal Outcomes in an Individual Randomised Controlled Trial." *Behaviour Research and Therapy* 130:103466.

McKelway, Madeline. 2020. "Women's Employment in India: Intra-Household and Intra-Personal Constraints." *Working Paper*.

McKenzie, David. 2012. "Beyond Baseline and Follow-up: The Case for More T in Experiments." *Journal of Development Economics* 99 (2): 210–221.

Ravensteijn, Bastian, and Eli Schachar. 2018. "The Impact of Cost Sharing for Mental Health Care on Mental Health Care Use and Labor Market Outcomes." Technical Report, Working Paper.

Ridley, Matthew W., Gautam Rao, Frank Schilbach, and Vikram H. Patel. 2020, May. "Poverty, Depression, and Anxiety: Causal Evidence and Mechanisms." Technical Report w27157, National Bureau of Economic Research.

Schoenbaum, Michael, Jürgen Unützer, Daniel McCaffrey, Naihua Duan, Cathy Sherbourne, and Kenneth B. Wells. 2002. "The Effects of Primary Care Depression Treatment on Patients' Clinical Status and Employment." *Health Services Research* 37 (5): 1145–1158.

Singla, Daisy R., Brandon A. Kohrt, Laura K. Murray, Arpita Anand, Bruce F. Chorpita, and Vikram Patel. 2017. "Psychological Treatments for the World: Lessons from Low- and Middle-Income Countries." *Annual Review of Clinical Psychology* 13 (May): 149–181.

World Health Organization. 2018. "Problem Management Plus (PM+): Individual Psychological Help for Adults Impaired by Distress in Communities Exposed to Adversity."

Figure 1: Design

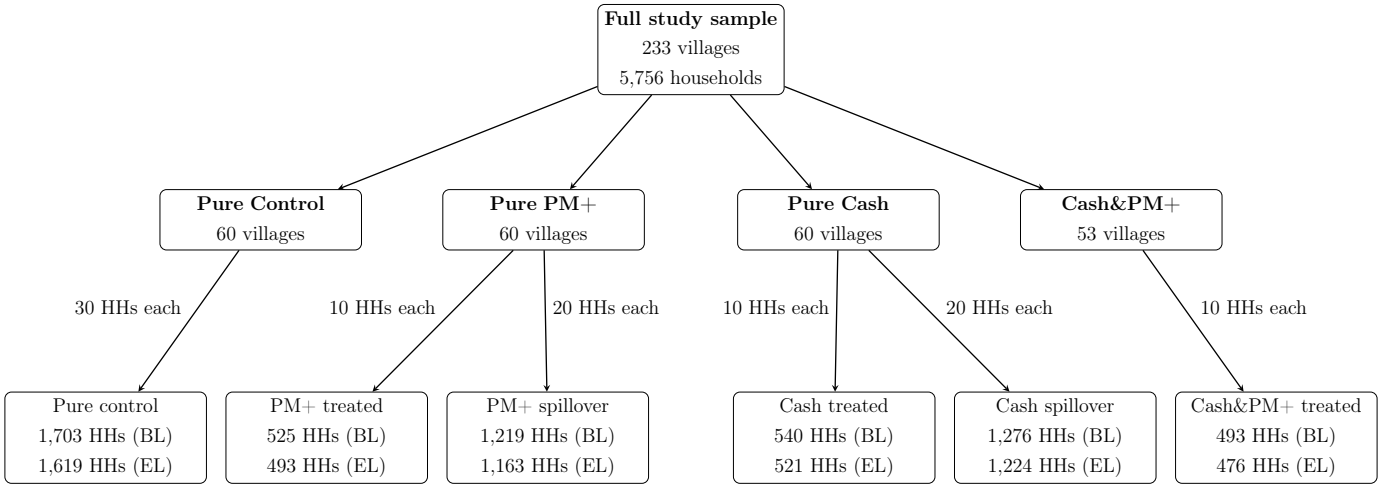
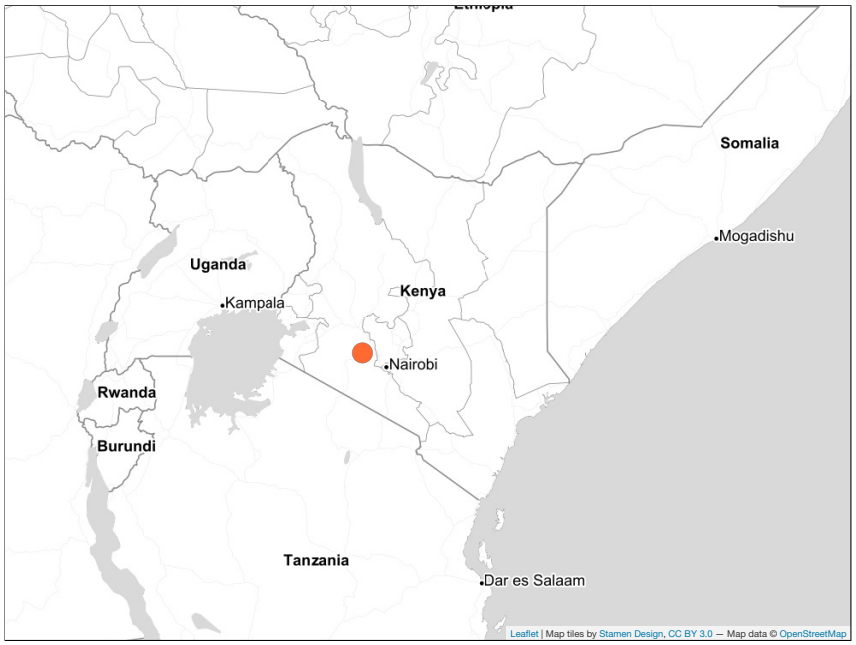
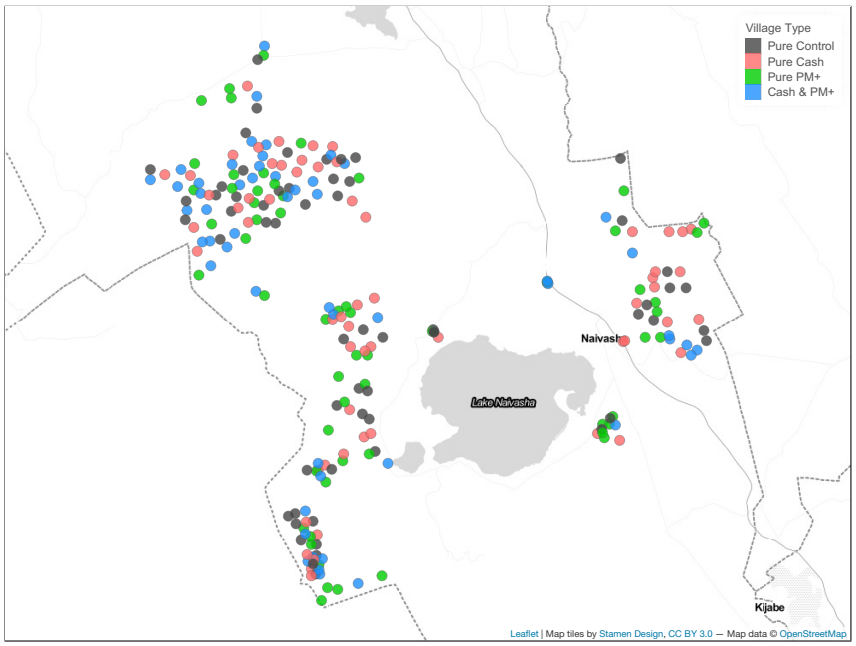


Figure 2: Study area

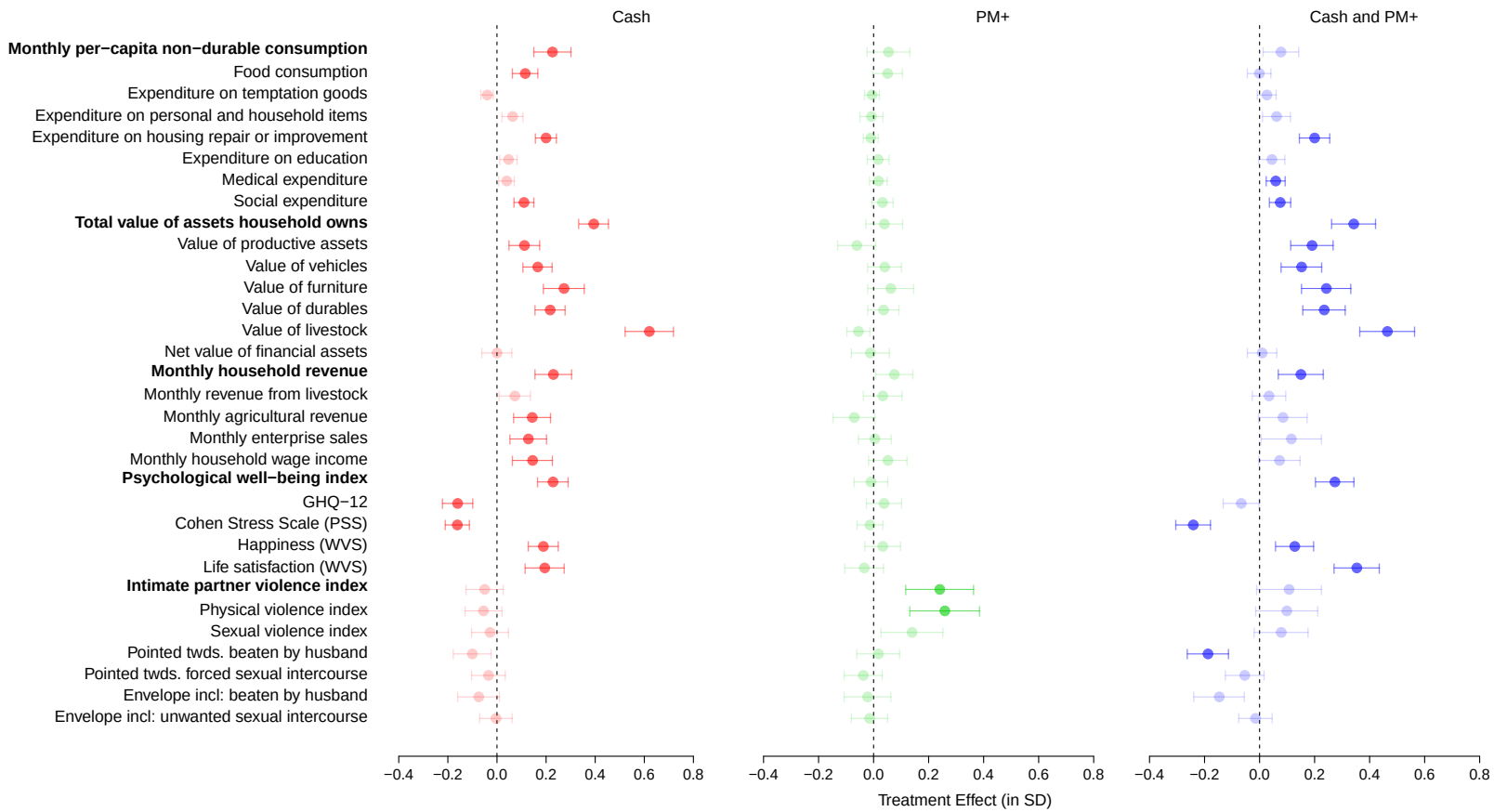


(a) Project area within Kenya



(b) Spatial distribution of treatment and control villages

Figure 3: Treatment effects



Notes: Estimated treatment coefficients for all primary outcomes and their immediate components. Each row represents a different outcome variable, with bolded rows denoting primary outcomes. To make outcomes comparable, we z-score all of them, so that effect sizes can be interpreted in standard deviation differences relative to to the pure control group mean. Each row displays three treatment effects, estimated by equation (1): pure cash (left), pure PM+ (middle), and the combined treatment (right). We also show one standard error around each point estimate. Estimates which are significantly different from zero at the 5 percent level are shown in darker shades. Note that for the psychological well-being index, lower values of GHQ-12 and Stress correspond to higher well-being.

Table 1: Treatment effects, primary outcomes

	(1) Pure control mean (SD)	(2) Cash treatment	(3) PM+ treatment	(4) Cash + PM+ treatment	(5) Cash = PM+	(6) Cash = Cash & PM+	(7) PM+ = Cash & PM+	(8) Cash & PM+ = Cash only + PM+ only	(9) N
Monthly per-capita non-durable consumption (USD PPP)	52.49	10.40	2.49	3.53	0.08*	0.09*	0.80	0.09*	5309
	(46.51)	(3.53)***	(3.61)	(3.00)					
		[0.00]***	[1.00]	[0.14]					
		+20%	+5%	+7%					
Total value of assets household owns (USD PPP)	553.16	272.16	28.75	236.07	0.00***	0.55	0.00***	0.39	5309
	(664.70)	(40.20)***	(44.97)	(53.68)***					
		[0.00]***	[1.00]	[0.00]***					
		+49%	+5%	+43%					
Monthly household revenue (USD PPP)	135.48	36.32	11.78	23.96	0.06*	0.40	0.38	0.18	5309
	(153.09)	(11.70)***	(10.54)	(12.47)*					
		[0.00]***	[1.00]	[0.06]*					
		+27%	+9%	+18%					
Psychological well-being index	0.00	0.24	−0.01	0.28	0.00***	0.56	0.00***	0.61	5309
	(1.00)	(0.06)***	(0.06)	(0.07)***					
		[0.00]***	[1.00]	[0.00]***					
Intimate partner violence index	−0.00	−0.05	0.25	0.08	0.02**	0.28	0.30	0.53	3052
	(1.00)	(0.08)	(0.12)**	(0.12)					
		[0.11]	[0.26]	[0.25]					

Notes: OLS estimates of treatment effects. Outcome variables, measured at endline, are listed on the left. The first column shows the pure control group mean and standard deviation (in parentheses) of the outcome variable. Columns (2), (3), and (4) show the treatment effects of the cash-only, PM+-only, and combined treatment groups, respectively, relative to the pure control group. Standard errors, clustered at the village level, are shown in parentheses, and FDR p-values in brackets. Columns (5)–(7) show p-values for the pairwise comparisons of these treatment arms. Column (8) shows p-values testing whether the sum of the cash-only and PM+-only treatment effects is the same as the treatment effect of the combined treatment. Column (9) shows the number of observations.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

Table 2: Treatment effects, secondary outcomes

	(1) Pure control mean (SD)	(2) Cash treatment	(3) PM+ treatment	(4) Cash + PM+ treatment	(5) Cash = PM+	(6) Cash = Cash & PM+	(7) PM+ = Cash & PM+	(8) Cash & PM+ = Cash only + PM+ only	(9) N
Food security index	0.00	0.14	−0.10	0.13	0.00***	0.82	0.00***	0.37	5309
	(1.00)	(0.06)**	(0.07)	(0.06)**					
Monthly household profits (USD PPP)	19.31	−4.71	−1.31	−6.88	0.58	0.77	0.42	0.92	5309
	(67.78)	(5.38)	(4.49)	(6.21)					
		–24%	–7%	–36%					
Working hours per week per capita	35.05	−0.50	−1.93	1.00	0.41	0.39	0.10*	0.14	5309
	(25.50)	(1.47)	(1.51)	(1.50)					
Education index	−0.00	−0.05	0.07	−0.07	0.18	0.85	0.16	0.48	1886
	(1.00)	(0.09)	(0.08)	(0.10)					

Notes: OLS estimates of treatment effects. Outcome variables, measured at endline, are listed on the left. The first column shows the pure control group mean and standard deviation (in parentheses) of the outcome variable. Columns (2), (3), and (4) show the treatment effects of the cash-only, PM+-only, and combined treatment groups, respectively, relative to the pure control group. Standard errors, clustered at the village level, are shown in parentheses. Columns (5)–(7) show p-values for the pairwise comparisons of these treatment arms. Column (8) shows p-values testing whether the sum of the cash-only and PM+-only treatment effects is the same as the treatment effect of the combined treatment. Column (9) shows the number of observations.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

Table 3: Treatment effects, well-being outcomes

	(1) Pure control mean (SD)	(2) Cash treatment	(3) PM+ treatment	(4) Cash + PM+ treatment	(5) Cash = PM+	(6) Cash = Cash & PM+	(7) PM+ = Cash & PM+	(8) Cash & PM+ = Cash only + PM+ only	(9) N
Psychological well-being index	0.00	0.24	−0.01	0.28	0.00***	0.56	0.00***	0.61	5309
	(1.00)	(0.06)***	(0.06)	(0.07)***					
GHQ-12	−0.00	−0.17	0.03	−0.07	0.00***	0.19	0.14	0.52	5309
	(1.00)	(0.06)***	(0.06)	(0.06)					
Perceived Stress Scale (Cohen)	−0.00	−0.17	−0.02	−0.24	0.00***	0.27	0.00***	0.46	5309
	(1.00)	(0.05)***	(0.05)	(0.06)***					
Happiness (WVS)	0.00	0.20	0.04	0.14	0.02**	0.42	0.20	0.31	5302
	(1.00)	(0.06)***	(0.07)	(0.07)**					
Life satisfaction (WVS)	0.00	0.23	−0.03	0.40	0.00***	0.08*	0.00***	0.10*	5309
	(1.00)	(0.08)***	(0.08)	(0.09)***					
WHODAS disability assessment	−0.00	−0.02	0.01	−0.00	0.64	0.80	0.84	0.96	5309
	(1.00)	(0.07)	(0.07)	(0.07)					

Notes: OLS estimates of treatment effects. Outcome variables, measured at endline, are listed on the left. The first column shows the pure control group mean and standard deviation (in parentheses) of the outcome variable. Columns (2), (3), and (4) show the treatment effects of the cash-only, PM+-only, and combined treatment groups, respectively, relative to the pure control group. Standard errors, clustered at the village level, are shown in parentheses. Columns (5)–(7) show *p*-values for the pairwise comparisons of these treatment arms. Column (8) shows *p*-values testing whether the sum of the cash-only and PM+-only treatment effects is the same as the treatment effect of the combined treatment. Column (9) shows the number of observations.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

For Online Publication: The Comparative Impact of Cash Transfers and Psychotherapy on Psychological and Economic Well-being

Johannes Haushofer, Robert Mudida, Jeremy Shapiro

Contents

A Integrity of the experiment **3**

A.1 Details on checks for data integrity 3

A.2 Timing 5

A.3 Baseline Balance 7

A.4 Attrition 8

B Direct treatment and spillover effects **9**

B.1 Direct treatment effects 9

B.2 Spillover effects 18

C Cost effectiveness **22**

D Heterogeneous impacts **23**

E CHW fixed effects **28**

F Treatment effects over time **29**

G Demand effects **30**

H Effects of transfer frequency **31**

I Indices and variables **33**

A. Integrity of the experiment

A.1 Details on checks for data integrity

1. High-frequency checks: These checks consisted of continuous monitoring of data coming into the server to check for missing observations and inconsistencies in responses. A STATA .do file was created and run regularly (at least weekly) on incoming data to check for errors. If any errors or discrepancies were detected, corrective action was taken to resolve these issues. Further, these checks informed the content of refresher training for field officers.

2. Back-checks: We performed back-checks with all household survey respondents on a subset of questions that elicited (mostly) immutable information (e.g. respondent's age, number of children). Back-checks were performed within a week of the initial survey by a different set of FOs, and the procedure was known to FOs *ex ante*. During baseline, 225 respondents were not reached on the second visit, primarily due to relocations. We included only respondents who were available for both the primary demographic survey and the back-check in the study to minimize later attrition.[1]

3. Random spot checks and field observations: Field officers were supervised by Project Leads, who regularly observed them while they conducted surveys. Specifically, project leads observed if questions were asked according to protocol, such as probing when answers are unclear. These checks were designed to ensure consistency of questioning across field officers. Continuous feedback was relayed to field officers on areas that needed improvement. Additionally, senior project management made random visits to the field.

4. GPS checks: GPS coordinates were recorded for all baseline and backcheck

[1] Respondents who refused to participate in the backcheck or who were estimated to be only temporarily unavailable by the FOs (37 respondents) were also included in the sample.

surveys. A different team member checked these coordinates on Google Earth to confirm the existence of a house at the specified location.

5. M-Pesa confirmation: for respondents receiving cash transfers, we confirmed that the M-Pesa numbers provided at baseline and backcheck matched, and that the name associated with the mobile money account matched the name of the intended recipient, before the transfer was initiated.

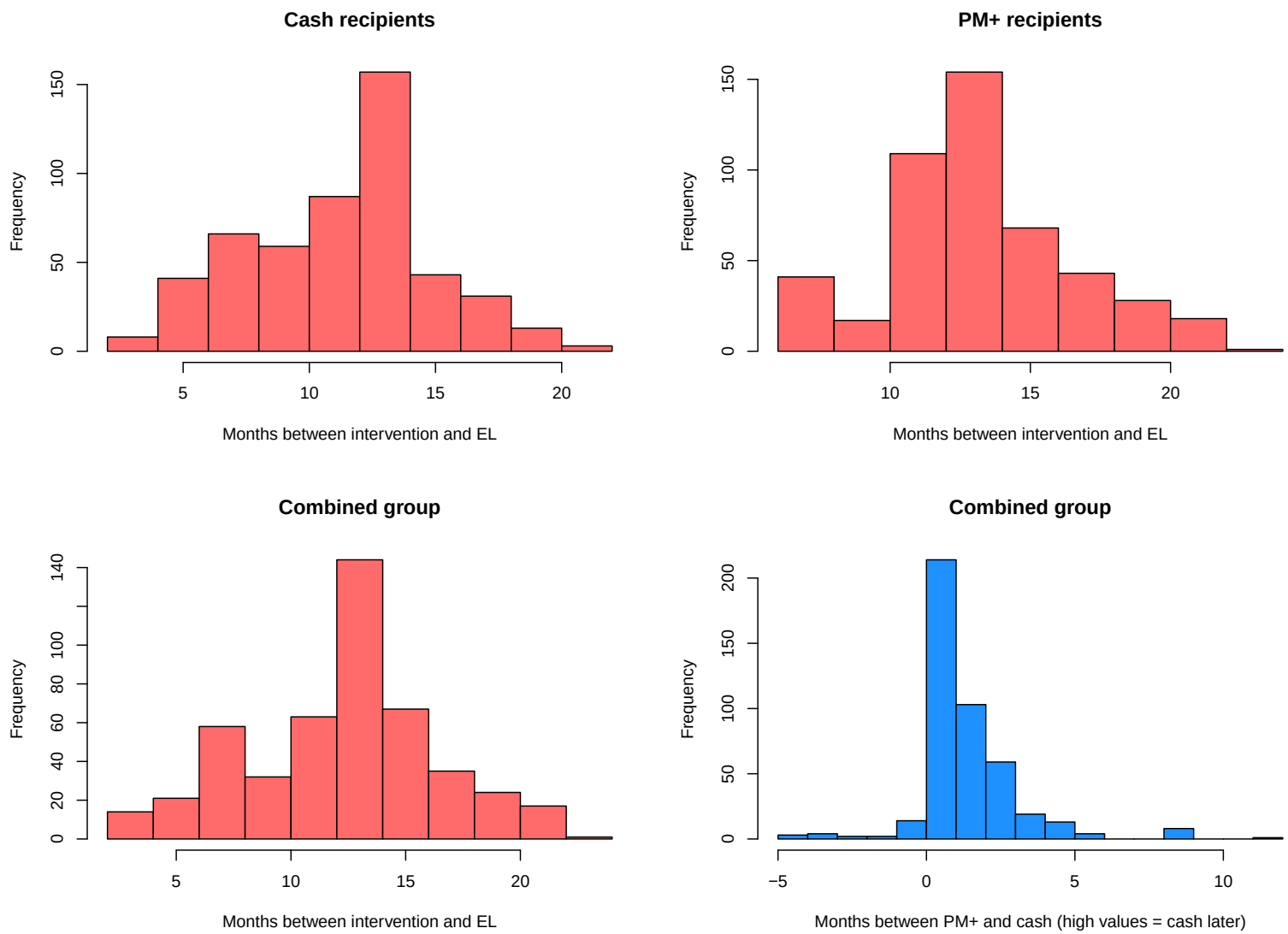
A.2 Timing

Table A.1: Timing

	(1) Mean	(2) Median	(3) Min	(4) Max	(5) N
Months between baseline and intervention start	9.91	10	2	19	1523
Months between intervention start and endline	13.95	14	3	28	1461
Months between intervention end and endline	12.61	13	2	23	1461
Cash recipients	11.99	13	3	22	977
PM+ recipients	13.52	14	2	23	934
Months between PM+ and cash (combined group)	1.36	1	-5	12	450
Months between baseline and endline	24.05	23	1	31	5492

Notes: Time in months between baseline, intervention start/end, and endline (mean, median, minimum, and maximum). The sample is restricted to the treatment group in the first three rows.

Figure A.1: Distribution of intervention timing



A.3 Baseline Balance

The table below shows OLS estimates of baseline balance. Outcome variables, measured at baseline, are listed on the left. The first column shows the pure control group mean and standard deviation (in parentheses) of the outcome variable. Columns (2), (3), (4), (5), and (6) show baseline differences between the pure control group and the cash-only, cash spillover, PM+-only, PM+ spillover, and Cash & PM+ groups, respectively. Standard errors, clustered at the village level, are shown in parentheses. Column (7) shows the number of observations.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

Table A.2: Baseline balance

	(1) Pure control mean (SD)	(2) Cash treatment	(3) Cash spillover	(4) PM+ treatment	(5) PM+ spillover	(6) Cash + PM+ treatment	(7) N
Female respondent	0.64	−0.00	−0.00	−0.00	−0.00	−0.00	5309
	(0.48)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	
Age	45.36	1.75	1.21	0.18	−0.60	0.93	5309
	(18.44)	(1.03)*	(0.87)	(1.14)	(0.88)	(1.15)	
Baseline asset index	0.00	0.03	0.00	0.01	0.02	−0.00	5309
	(1.00)	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)	
Baseline M-Pesa access	0.80	0.00	−0.00	0.00	−0.00	0.00	5309
	(0.40)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	
Baseline psychological well-being index	0.00	0.05	0.03	0.02	0.06	0.03	5309
	(1.00)	(0.04)	(0.03)	(0.04)	(0.03)*	(0.04)	
Baseline psychological distress index	0.00	−0.01	−0.03	−0.00	0.02	0.06	5309
	(1.00)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	
Baseline IPV index	−0.00	−0.01	−0.01	0.02	0.03	0.12	3212
	(1.00)	(0.07)	(0.06)	(0.08)	(0.07)	(0.13)	
Baseline justifiability of violence	−0.00	0.18	−0.13	0.07	1.00	0.15	3212
	(1.00)	(0.18)	(0.07)*	(0.09)	(0.92)	(0.11)	

A.4 Attrition

The table below shows OLS estimates of attrition between baseline and endline. Column 1 regresses an indicator for being observed at both baseline and endline on treatment group assignment across all villages. Standard errors, clustered at the village level, are shown in parentheses. Columns (2) and (3) restrict the sample to pure PM+ and pure cash villages, respectively, and include village-level fixed effects, to assess attrition between treatment and spillover households in those villages. Unclustered standard errors are shown in parentheses.

Table A.3: Attrition

	All villages	Pure PM+ villages	Pure cash Villages
	(1)	(2)	(3)
	$\mathbb{1}$(Completed both surveys)	$\mathbb{1}$(Completed both surveys)	$\mathbb{1}$(Completed both surveys)
Pure PM+	−0.00		
	(0.01)		
Pure cash	0.01		
	(0.01)		
Cash + PM+	0.01		
	(0.02)		
Treatment		−0.02	−0.01
		(0.01)	(0.01)
Pure control village mean	0.95		
Control mean		0.95	0.97
N	5756	1744	1777

B. Direct treatment and spillover effects

B.1 Direct treatment effects

The tables below show OLS estimates of direct treatment effects. Outcome variables, measured at endline, are listed on the left. The first column shows the pure control group mean and standard deviation (in parentheses) of the outcome variable. Columns (2), (3), and (4) show the treatment effects of the cash-only, PM+-only, and combined treatment groups, respectively, relative to the pure control group. Standard errors, clustered at the village level, are shown in parentheses, and FDR p-values in brackets where applicable. Columns (5)–(7) show p-values for the pairwise comparisons of these treatment arms. Column (8) shows p-values testing whether the sum of the cash-only and PM+-only treatment effects is the same as the treatment effect of the combined treatment. Column (9) shows the number of observations.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

Table B.1: Treatment effects, asset outcomes

	(1) Pure control mean (SD)	(2) Cash treatment	(3) PM+ treatment	(4) Cash + PM+ treatment	(5) Cash = PM+	(6) Cash = Cash & PM+	(7) PM+ = Cash & PM+	(8) Cash & PM+ = Cash only + PM+ only	(9) N
Total value of assets household owns (USD PPP)	553.16	272.16	28.75	236.07	0.00***	0.55	0.00***	0.39	5309
	(664.70)	(40.20)***	(44.97)	(53.68)***					
		+49%	+5%	+43%					
Value of productive assets (USD PPP)	37.04	6.92	−3.17	11.17	0.03**	0.37	0.01***	0.23	5309
	(55.36)	(3.52)*	(3.93)	(4.31)**					
		+19%	–9%	+30%					
Value of vehicles (USD PPP)	114.76	60.34	15.00	55.24	0.08*	0.86	0.17	0.58	5309
	(342.36)	(20.34)***	(21.17)	(25.21)**					
		+53%	+13%	+48%					
Value of furniture (USD PPP)	288.12	74.79	17.40	66.25	0.02**	0.75	0.07*	0.45	5309
	(258.02)	(21.53)***	(21.30)	(22.91)***					
		+26%	+6%	+23%					
Value of durables (USD PPP)	122.65	36.31	6.31	38.81	0.01***	0.86	0.02**	0.82	5309
	(157.00)	(9.34)***	(8.83)	(12.23)***					
		+30%	+5%	+32%					
Value of livestock (USD PPP)	32.70	76.99	−6.70	57.69	0.00***	0.25	0.00***	0.47	5309
	(123.21)	(12.16)***	(5.19)	(12.35)***					
		+235%	–20%	+176%					
Net value of financial assets (USD PPP)	−53.86	−1.48	−3.14	0.81	0.93	0.88	0.82	0.80	5309
	(214.71)	(13.23)	(14.73)	(11.59)					
		–3%	–6%	+1%					

Table B.2: Treatment effects, consumption outcomes

	(1) Pure control mean (SD)	(2) Cash treatment	(3) PM+ treatment	(4) Cash + PM+ treatment	(5) Cash = PM+	(6) Cash = Cash & PM+	(7) PM+ = Cash & PM+	(8) Cash & PM+ = Cash only + PM+ only	(9) N
Monthly per-capita non-durable consumption (USD PPP)	52.49	10.40	2.49	3.53	0.08*	0.09*	0.80	0.09*	5309
	(46.51)	(3.53)***	(3.61)	(3.00)					
		+20%	+5%	+7%					
Food consumption (USD PPP)	48.91	5.93	2.65	−0.18	0.35	0.05**	0.37	0.04**	5309
	(53.01)	(2.70)**	(2.84)	(2.27)					
		+12%	+5%	−0%					
Expenditure on temptation goods (USD PPP)	0.60	−0.08	−0.01	0.05	0.25	0.08*	0.44	0.11	5309
	(1.88)	(0.05)*	(0.05)	(0.07)					
		−13%	−2%	+8%					
Expenditure on personal and household items (USD PPP)	6.14	0.54	−0.06	0.52	0.17	0.98	0.24	0.94	5309
	(8.57)	(0.37)	(0.35)	(0.44)					
		+9%	−1%	+9%					
Expenditure on housing repair or improvement (USD PPP)	2.85	1.92	−0.09	1.92	0.00***	1.00	0.00***	0.89	5309
	(9.57)	(0.41)***	(0.26)	(0.53)***					
		+67%	−3%	+67%					
Expenditure on education (USD PPP)	6.63	0.59	0.21	0.56	0.44	0.95	0.57	0.73	5309
	(11.13)	(0.40)	(0.44)	(0.53)					
		+9%	+3%	+8%					
Medical expenditure (USD PPP)	3.62	0.34	0.16	0.51	0.56	0.63	0.31	0.98	5309
	(8.64)	(0.27)	(0.28)	(0.30)*					
		+10%	+4%	+14%					
Social expenditure (USD PPP)	1.43	0.31	0.09	0.21	0.11	0.48	0.36	0.28	5309
	(2.77)	(0.11)***	(0.11)	(0.11)*					
		+22%	+6%	+15%					

Table B.3: Treatment effects, revenue outcomes

	(1) Pure control mean (SD)	(2) Cash treatment	(3) PM+ treatment	(4) Cash + PM+ treatment	(5) Cash = PM+	(6) Cash = Cash & PM+	(7) PM+ = Cash & PM+	(8) Cash & PM+ = Cash only + PM+ only	(9) N
Monthly household revenue (USD PPP)	135.48	36.32	11.78	23.96	0.06*	0.40	0.38	0.18	5309
	(153.09)	(11.70)***	(10.54)	(12.47)*					
		+27%	+9%	+18%					
Monthly revenue from livestock (USD PPP)	29.42	5.23	2.20	2.88	0.52	0.59	0.88	0.44	5309
	(56.58)	(3.72)	(4.08)	(3.60)					
		+18%	+8%	+10%					
Monthly agricultural revenue (USD PPP)	14.12	2.79	−1.17	1.71	0.01***	0.54	0.10*	0.97	5309
	(17.72)	(1.35)**	(1.35)	(1.59)					
		+20%	−8%	+12%					
Monthly enterprise sales (USD PPP)	9.68	6.64	0.24	6.00	0.14	0.92	0.34	0.90	5309
	(51.83)	(3.85)*	(3.08)	(5.67)					
		+69%	+3%	+62%					
Monthly household wage income (USD PPP)	78.01	16.10	5.71	7.97	0.29	0.43	0.81	0.28	5309
	(112.98)	(9.14)*	(7.80)	(8.45)					
		+21%	+7%	+10%					

Table B.4: Treatment effects, IPV outcomes

	(1) Pure control mean (SD)	(2) Cash treatment	(3) PM+ treatment	(4) Cash + PM+ treatment	(5) Cash = PM+	(6) Cash = Cash & PM+	(7) PM+ = Cash & PM+	(8) Cash & PM+ = Cash only + PM+ only	(9) N
Intimate partner violence index	−0.00	−0.05	0.25	0.08	0.02**	0.28	0.30	0.53	3052
	(1.00)	(0.08)	(0.12)**	(0.12)					
Physical violence index	0.00	−0.06	0.27	0.08	0.01***	0.21	0.23	0.49	3052
	(1.00)	(0.07)	(0.12)**	(0.11)					
Sexual violence index	0.00	−0.03	0.14	0.06	0.15	0.40	0.58	0.76	3045
	(1.00)	(0.07)	(0.11)	(0.10)					
Emotional violence index	0.00	0.11	0.09	−0.01	0.88	0.32	0.36	0.16	3052
	(1.00)	(0.11)	(0.09)	(0.09)					
Male attitudes index	0.00	−0.09	−0.01	−0.05	0.42	0.73	0.69	0.72	3052
	(1.00)	(0.08)	(0.08)	(0.09)					
Justifiability of violence index	0.00	−0.03	0.11	0.11	0.21	0.21	0.97	0.86	3052
	(1.00)	(0.09)	(0.10)	(0.09)					
Satisfaction with relationship index	0.00	0.07	−0.15	−0.13	0.01***	0.02**	0.81	0.67	3051
	(1.00)	(0.07)	(0.07)**	(0.08)					
Violence against children indicator	0.63	0.05	0.05	0.04	0.96	0.84	0.81	0.26	3212
	(0.48)	(0.03)	(0.04)	(0.03)					

Table B.5: Treatment effects, IPV further outcomes

	(1) Pure control mean (SD)	(2) Cash treatment	(3) PM+ treatment	(4) Cash + PM+ treatment	(5) Cash = PM+	(6) Cash = Cash & PM+	(7) PM+ = Cash & PM+	(8) Cash & PM+ = Cash only + PM+ only	(9) N
Intimate partner violence index	−0.00	−0.05	0.25	0.08	0.02**	0.28	0.30	0.53	3052
	(1.00)	(0.08)	(0.12)**	(0.12)					
Envelope incl: beaten by husband	0.74	−0.03	−0.00	−0.06	0.59	0.39	0.17	0.57	3212
	(0.44)	(0.04)	(0.04)	(0.04)					
Envelope incl: unwanted sexual intercourse	0.56	−0.00	−0.00	−0.01	0.94	0.85	0.91	0.96	3212
	(0.50)	(0.03)	(0.03)	(0.03)					
Pointed twds. beaten by husband	0.59	−0.05	0.01	−0.09	0.16	0.31	0.01**	0.39	3212
	(0.49)	(0.04)	(0.04)	(0.04)**					
Pointed twds. forced sexual intercourse	0.54	−0.02	−0.02	−0.03	0.98	0.82	0.80	0.85	3212
	(0.50)	(0.03)	(0.03)	(0.03)					

Table B.6: Treatment effects, food security outcomes

	(1) Pure control mean (SD)	(2) Cash treatment	(3) PM+ treatment	(4) Cash + PM+ treatment	(5) Cash = PM+	(6) Cash = Cash & PM+	(7) PM+ = Cash & PM+	(8) Cash & PM+ = Cash only + PM+ only	(9) N
Food security index	0.00	0.14	−0.10	0.13	0.00***	0.82	0.00***	0.37	5309
	(1.00)	(0.06)**	(0.07)	(0.06)**					
Number of times last month adults cut or skipped meals	1.82	−0.21	0.34	−0.11	0.01**	0.60	0.04**	0.37	5309
	(4.79)	(0.15)	(0.19)*	(0.16)					
Number of times last month children cut or skipped meals	0.77	−0.10	0.17	0.00	0.01***	0.17	0.10*	0.55	5309
	(2.37)	(0.06)*	(0.09)*	(0.07)					
Number of times last month had to borrow food	0.95	−0.10	0.06	0.04	0.17	0.17	0.89	0.59	5309
	(2.55)	(0.09)	(0.11)	(0.10)					
All household members regularly eat at least twice a day	0.91	0.03	0.00	0.03	0.13	0.89	0.20	0.88	5309
	(0.28)	(0.01)**	(0.02)	(0.02)*					
All household members regularly eat until they are content	0.87	0.05	0.03	0.04	0.41	0.71	0.67	0.19	5309
	(0.34)	(0.02)***	(0.02)	(0.02)**					
Number of times last week respondent has eggs, meat or fish	1.11	0.12	−0.09	0.16	0.02**	0.62	0.00***	0.25	5309
	(1.43)	(0.09)	(0.08)	(0.07)**					

Table B.7: Treatment effects, labor supply outcomes

	(1) Pure control mean (SD)	(2) Cash treatment	(3) PM+ treatment	(4) Cash + PM+ treatment	(5) Cash = PM+	(6) Cash = Cash & PM+	(7) PM+ = Cash & PM+	(8) Cash & PM+ = Cash only + PM+ only	(9) N
Working hours per week per capita	35.05	−0.50	−1.93	1.00	0.41	0.39	0.10*	0.14	5309
	(25.50)	(1.47)	(1.51)	(1.50)					
Hours working in agriculture for this household	16.04	−0.77	−2.24	0.71	0.21	0.21	0.02**	0.02**	5309
	(16.84)	(1.09)	(1.14)*	(1.15)					
Hours tending animals for this household	10.07	0.80	−0.38	1.83	0.15	0.24	0.01**	0.22	5309
	(15.24)	(0.77)	(0.76)	(0.83)**					
Hours working in non-farm household business	4.08	0.42	1.09	1.19	0.34	0.24	0.89	0.73	5309
	(11.57)	(0.47)	(0.64)*	(0.58)**					
Hours working for pay for someone outside this household	15.50	−0.95	−0.51	−2.66	0.72	0.12	0.06*	0.45	5309
	(19.88)	(1.14)	(1.16)	(1.03)**					

Table B.8: Treatment effects, education outcomes

	(1) Pure control mean (SD)	(2) Cash treatment	(3) PM+ treatment	(4) Cash + PM+ treatment	(5) Cash = PM+	(6) Cash = Cash & PM+	(7) PM+ = Cash & PM+	(8) Cash & PM+ = Cash only + PM+ only	(9) N
Education index	−0.00	−0.05	0.07	−0.07	0.18	0.85	0.16	0.48	1886
	(1.00)	(0.09)	(0.08)	(0.10)					
Proportion of children (< 19) in school	0.93	−0.00	−0.01	0.00	0.96	0.45	0.47	0.40	5309
	(0.19)	(0.01)	(0.01)	(0.01)					
Days of school missed per child in last 30 days	9.88	−0.11	0.10	−0.18	0.61	0.86	0.50	0.74	5309
	(6.04)	(0.33)	(0.33)	(0.33)					
Annual expenditure on school expenses per child (USD PPP)	807.52	4.11	19.15	−13.22	0.63	0.58	0.32	0.37	5309
	(466.12)	(24.17)	(24.96)	(26.38)					
Time spent studying per child in last week	68.57	−1.39	0.97	−1.29	0.18	0.96	0.21	0.71	5309
	(24.40)	(1.56)	(1.30)	(1.63)					

B.2 Spillover effects

These tables show OLS estimates of spillover effects. Outcome variables, measured at endline, are listed on the left. The first column shows the pure control group mean and standard deviation (in parentheses) of the outcome variable. Columns (2) and (3) show the spillover effects of the cash-only and PM+-only treatments, respectively, relative to the pure control group. Standard errors, clustered at the village level, are shown in parentheses, and FDR p-values in brackets, where applicable. Column (4) shows p-values for the pairwise comparison of these two spillover effects. Column (5) shows the number of observations.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

Table B.9: Spillover effects, primary outcomes

	(1) Pure control mean (SD)	(2) Cash spillover	(3) PM+ spillover	(4) Cash = PM+	(5) N
Monthly per-capita non-durable consumption (USD PPP)	52.49	3.46	1.42	0.45	5309
	(46.51)	(2.46)	(2.52)		
		[0.68]	[0.75]		
		+7%	+3%		
Total value of assets household owns (USD PPP)	553.16	13.89	29.29	0.68	5309
	(664.70)	(35.23)	(33.19)		
		[1.00]	[0.61]		
		+3%	+5%		
Monthly household revenue (USD PPP)	135.48	0.36	22.00	0.03**	5309
	(153.09)	(8.53)	(10.28)**		
		[1.00]	[0.20]		
		+0%	+16%		
Psychological well-being index	0.00	0.02	0.02	0.93	5309
	(1.00)	(0.05)	(0.06)		
		[1.00]	[0.87]		
Intimate partner violence index	−0.00	0.17	0.17	1.00	3052
	(1.00)	(0.10)*	(0.09)*		
		[0.68]	[0.20]		

Table B.10: Spillover effects, secondary outcomes

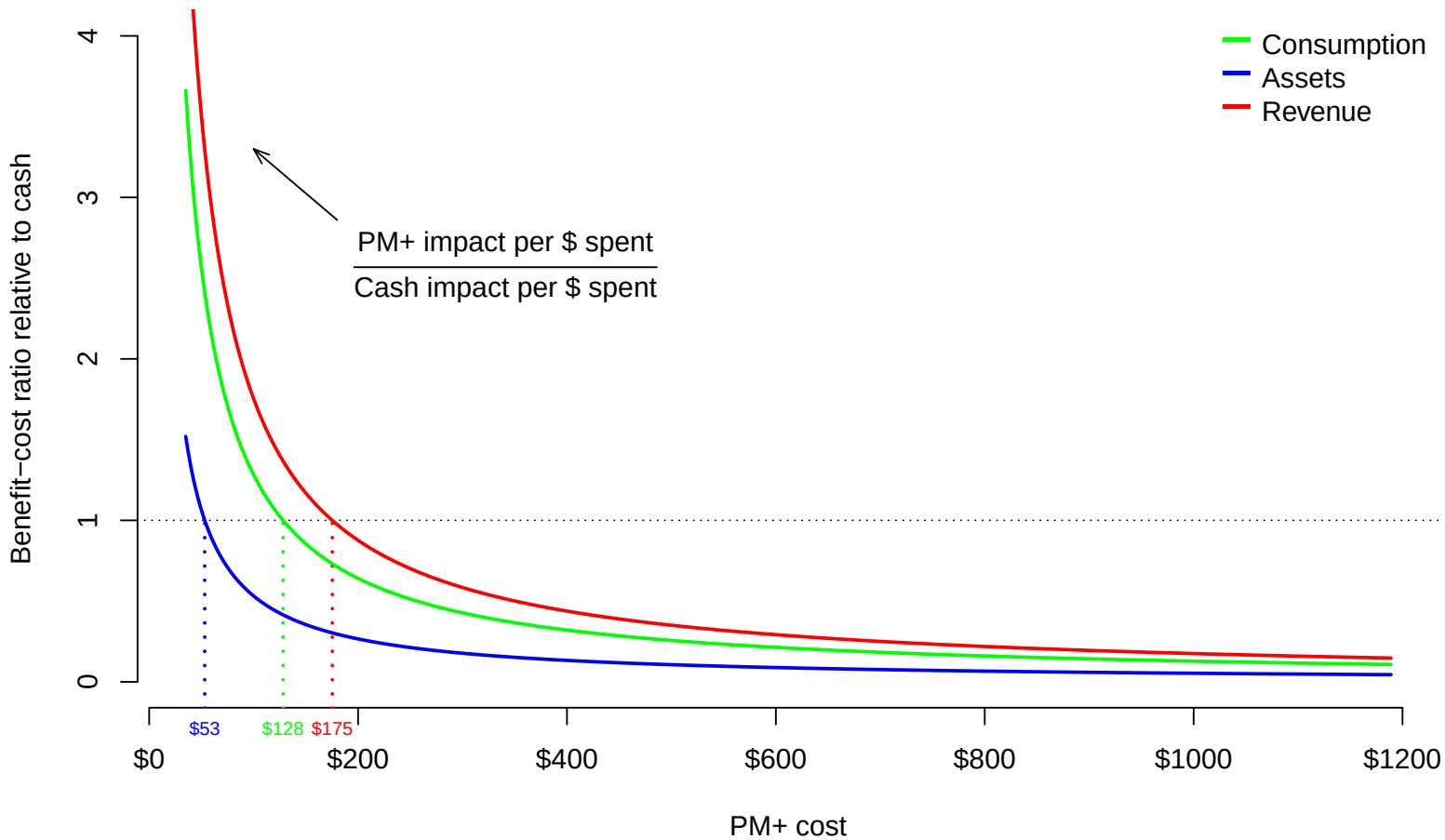
	(1) Pure control mean (SD)	(2) Cash spillover	(3) PM+ spillover	(4) Cash = PM+	(5) N
Food security index	0.00	−0.05	−0.07	0.75	5309
	(1.00)	(0.06)	(0.06)		
Monthly household profits (USD PPP)	19.31	−2.20	1.05	0.36	5309
	(67.78)	(3.20)	(3.80)		
		−11%	+5%		
Working hours per week per capita	35.05	−0.80	−0.96	0.91	5309
	(25.50)	(1.42)	(1.26)		
Education index	−0.00	−0.15	0.03	0.07*	1886
	(1.00)	(0.09)*	(0.08)		

Table B.11: Spillover effects, well-being outcomes

	(1) Pure control mean (SD)	(2) Cash spillover	(3) PM+ spillover	(4) Cash = PM+	(5) N
Psychological well-being index	0.00	0.02	0.02	0.93	5309
	(1.00)	(0.05)	(0.06)		
GHQ-12	−0.00	−0.03	−0.04	0.79	5309
	(1.00)	(0.06)	(0.05)		
Perceived Stress Scale (Cohen)	−0.00	−0.00	−0.01	0.85	5309
	(1.00)	(0.04)	(0.05)		
Happiness (WVS)	0.00	−0.02	0.02	0.56	5302
	(1.00)	(0.06)	(0.06)		
Life satisfaction (WVS)	0.00	0.05	−0.01	0.43	5309
	(1.00)	(0.07)	(0.07)		
WHODAS disability assessment	−0.00	−0.01	−0.09	0.12	5309
	(1.00)	(0.06)	(0.06)		

C. Cost effectiveness

Figure C.1: Cost effectiveness



Notes: Benefit-cost ratio of PM+ relative to cash transfers. The horizontal axis reflects different potential cost of delivering the PM+ treatment to one person, ranging from marginal cost (USD 37 nominal per person) to total cost (USD 1,189 nominal per person) in this study. The vertical axis plots the ratio of the impact of the PM+ (only) treatment on monetary outcomes, per dollar spent on PM+, relative to the impact of the cash transfer (only) treatment on the same outcomes, per dollar spent on cash transfers. To the left of the dotted vertical lines, PM+ has a larger impact per dollar spent than cash transfers on the outcome variable in question.

D. Heterogeneous impacts

To explore whether our estimated treatment effects vary over a set of pre-specified heterogeneity dimensions, we re-estimate our main equation with an additional interaction term on one of the five main treatment or spillover dummies CT_{vi}, PMP_{vi}, CT/PMP_{vi}, $SPILLCT_{vi}$ and $SPILLPMP_{vi}$. For example, to investigate whether the effect of the cash treatment is differential for recipients with high baseline assets ($HA_{vi} = 1$), we estimate

$$\begin{aligned} y_{vi} = \alpha + \beta_1 CT_{vi} + \beta_h CT_{vi} \times HA_{vi} + \gamma_h HA_{vi} + \beta_2 PMP_{vi} + \beta_3 CT\&PMP_{vi} \\ + \beta_4 DFS_{vi} + \beta_5 SpillCT_{iv} + \beta_6 SpillPMP + \gamma X_{vi} + \delta y_{viB} + \varepsilon_{vi} \end{aligned} \tag{2}$$

and report β_h. We repeat this exercise for each of the treatment groups, and for each of the pre-specified heterogeneity dimensions: a dummy for holding higher-than-median baseline assets ("high assets"), reporting higher-than-median baseline intimate partner violence ("high IPV"), being female, or reporting higher-than-median psychological distress ("high distress") at baseline. The tables below show OLS estimates of these analyses. Outcome variables, measured at endline, are listed on the left. The first column shows the pure control group mean and standard deviation (in parentheses) of the outcome variable. Columns (1)–(5) show the interaction term in the heterogeneous treatment effects specifications for four interactants: asset holdings at baseline (column 1; median split); IPV at baseline (column 2; median split); gender (column 3); and high psychological distress at baseline (column 4: 33 percent most distressed; column 5: median split). Each coefficient can be thought of as the difference in treatment effects in the group that scores high on the interactant to the group that scores low on the interactant. Note that all coefficients in the table hence stem from different estimations. Standard errors, clustered at the village level, are shown in parentheses.

* p $<$ 0.1, ** p $<$ 0.05, *** p $<$ 0.01. Standard errors in parentheses.

Table D.1: Heterogeneous cash treatment effects, primary outcomes

	(1) Cash treatment × high assets	(2) Cash treatment × high IPV	(3) Cash treatment × female	(4) Cash treatment × high distress (top 33%)	(5) Cash treatment × high distress (top 50%)
Monthly per-capita non-durable consumption (USD PPP)	−6.47	−3.19	−13.36	−8.41	−1.49
	(4.60)	(5.54)	(6.00)**	(5.41)	(5.23)
Total value of assets household owns (USD PPP)	127.93	−163.42	−91.95	−48.42	0.10
	(66.91)*	(105.72)	(82.28)	(75.53)	(64.54)
Monthly household revenue (USD PPP)	7.80	−17.95	18.91	−36.70	−13.97
	(17.33)	(22.24)	(16.96)	(14.50)**	(16.38)
Psychological well-being index	0.00	−0.03	0.06	0.05	0.12
	(0.09)	(0.11)	(0.10)	(0.09)	(0.09)
Intimate partner violence index	0.28	0.02	−0.10	0.06	0.14
	(0.12)**	(0.11)	(0.13)	(0.14)	(0.13)

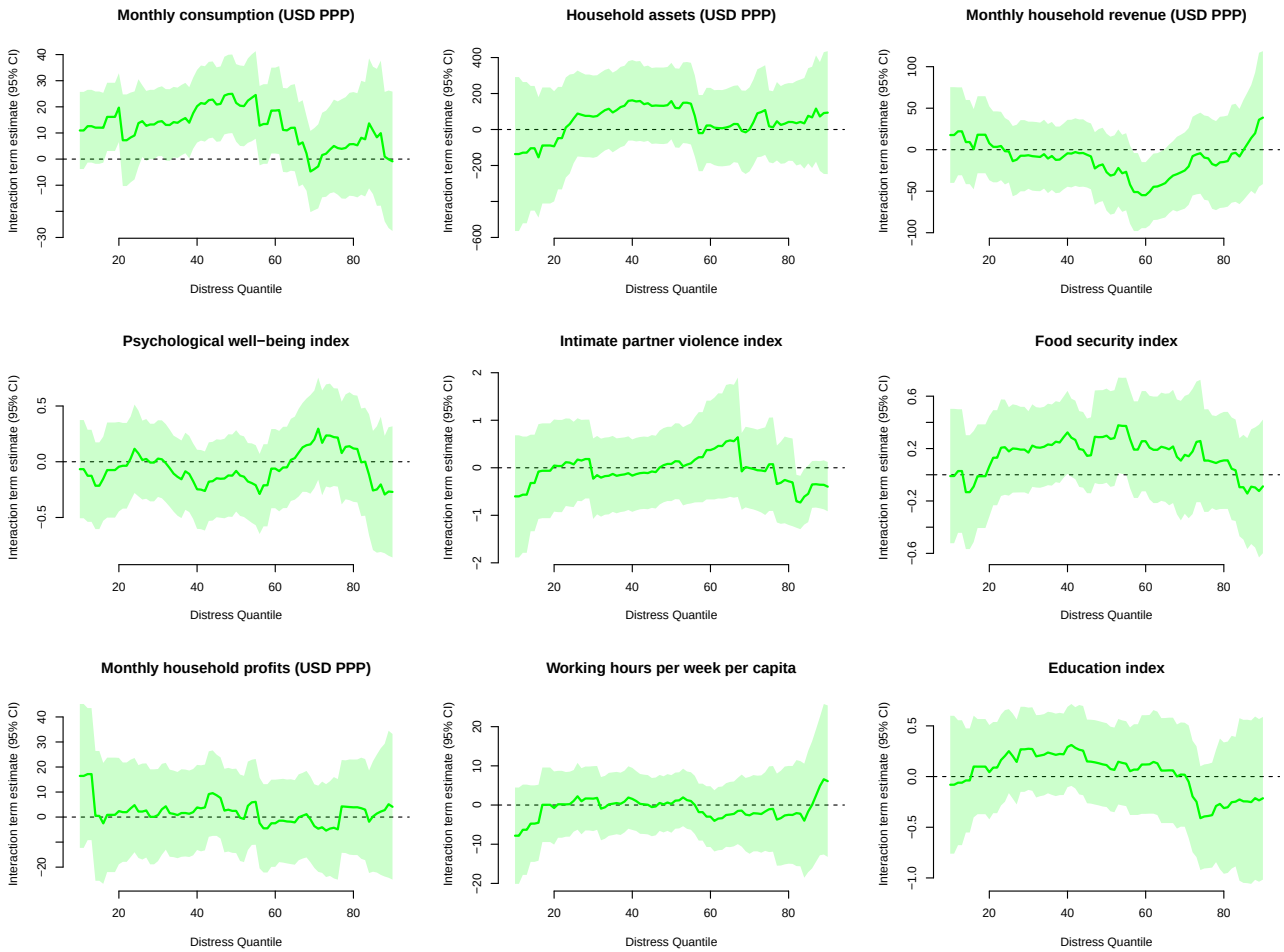
Table D.2: Heterogeneous PM+ treatment effects, primary outcomes

	(1) PM+ treatment × high assets	(2) PM+ treatment × high IPV	(3) PM+ treatment × female	(4) PM+ treatment × high distress (top 33%)	(5) PM+ treatment × high distress (top 50%)
Monthly per-capita non-durable consumption (USD PPP)	4.27	−5.13	0.89	1.49	2.02
	(5.66)	(5.68)	(6.89)	(4.90)	(5.22)
Total value of assets household owns (USD PPP)	−77.85	−28.91	85.94	44.27	−18.28
	(65.70)	(85.36)	(82.72)	(67.01)	(57.89)
Monthly household revenue (USD PPP)	1.37	8.31	13.07	12.77	11.52
	(15.70)	(18.59)	(16.44)	(15.96)	(17.00)
Psychological well-being index	−0.11	−0.16	−0.13	0.05	0.06
	(0.10)	(0.12)	(0.11)	(0.11)	(0.09)
Intimate partner violence index	−0.19	0.19	0.05	0.06	0.06
	(0.24)	(0.24)	(0.21)	(0.25)	(0.19)

Table D.3: Heterogeneous cash & PM+ treatment effects, primary outcomes

	(1) Cash & PM+ treatment × high assets	(2) Cash & PM+ treatment × high IPV	(3) Cash & PM+ treatment × female	(4) Cash & PM+ treatment × high distress (top 33%)	(5) Cash & PM+ treatment × high distress (top 50%)
Monthly per-capita non-durable consumption (USD PPP)	5.57	7.97	−2.09	−2.59	1.92
	(5.06)	(4.57)*	(5.05)	(4.32)	(4.58)
Total value of assets household owns (USD PPP)	106.46	−16.79	132.35	98.33	22.01
	(80.28)	(127.48)	(77.81)*	(87.22)	(85.04)
Monthly household revenue (USD PPP)	2.13	39.34	3.31	1.27	−8.67
	(18.87)	(19.77)**	(20.29)	(15.86)	(16.82)
Psychological well-being index	0.06	0.07	0.05	0.02	0.06
	(0.09)	(0.12)	(0.10)	(0.11)	(0.10)
Intimate partner violence index	−0.31	−0.18	−0.23	−0.25	−0.04
	(0.28)	(0.17)	(0.21)	(0.19)	(0.16)

Figure D.1: Heterogeneous effects of PM+ on severely distressed participants

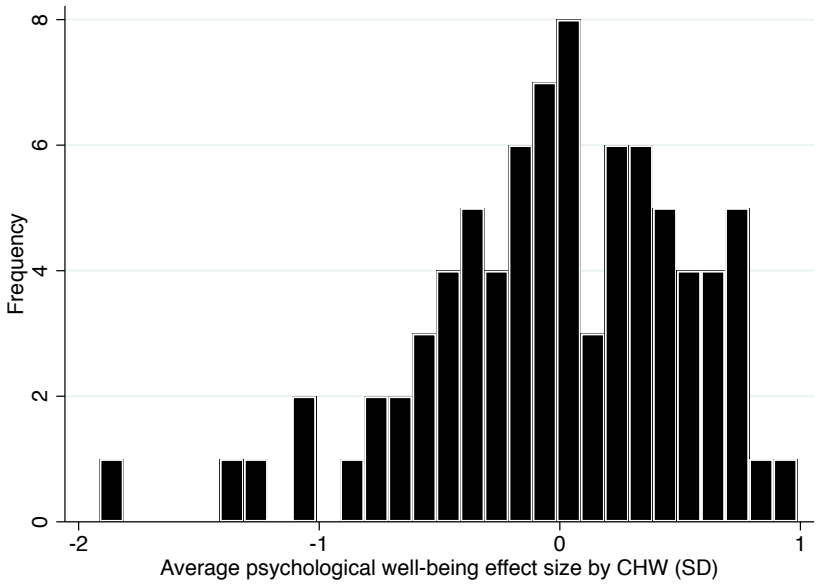


Notes: Estimates for additional heterogeneity analyses of PM+ on the subsample of participants classified as highly distressed (top 33%) at baseline. Within this subgroup, heterogeneous effects estimations are conducted for the sub-sub-sample of participants that are considered *severely distressed* based on a baseline distress index (weighted index of GHQ and WHODAS scores). The definition of severe distress gradually becomes more stringent from left to right along the x-axis, beginning with the entire 33% on the left and increasing the cutoff until only one person is considered distressed. The panels plot the interaction coefficients and 95% confidence bands for primary and secondary outcomes.

E. CHW fixed effects

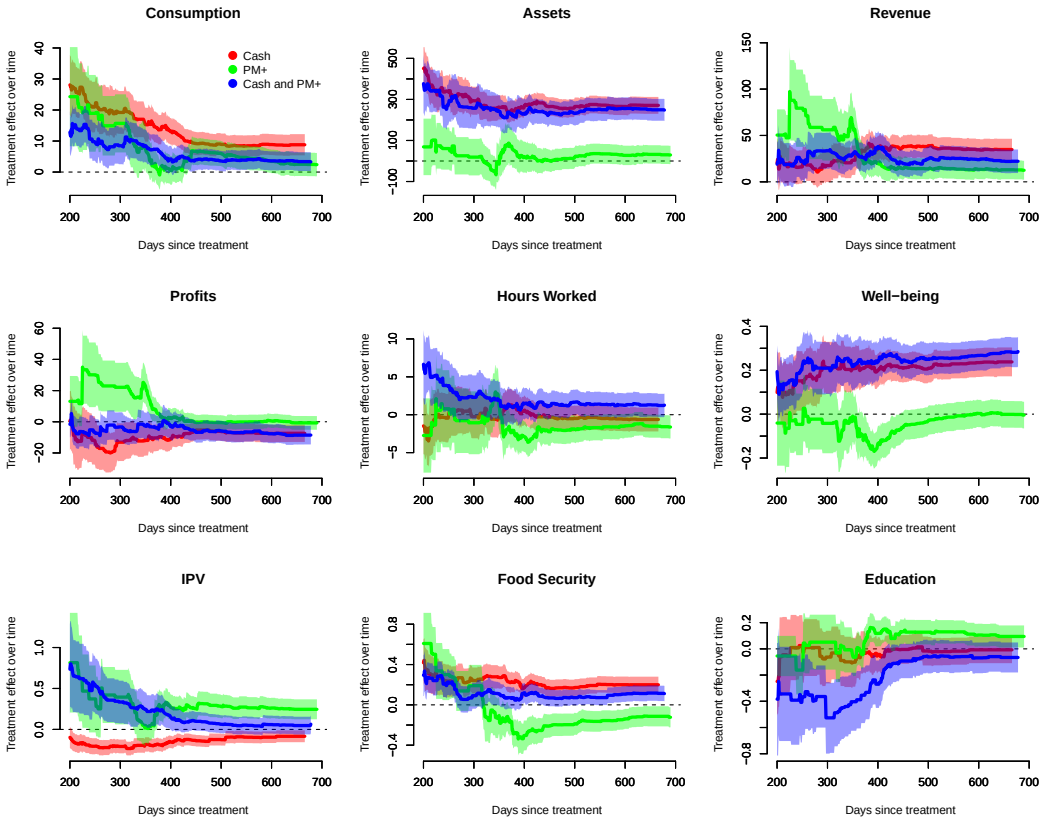
To understand whether the impact of PM+ on psychological well-being differed by which of the CHWs delivered the intervention, we restrict the sample to the PM+-only and pure control groups, and regress the z-scored psychological well-being index on a set of dummies for the CHWs. The histogram below shows the distribution of treatment effects by CHW.

Figure E.1: Distribution of PM+ treatment effects on psychological well-being by CHW



F. Treatment effects over time

Figure F.1: Treatment effects over time



Notes: Treatment effect estimates using our main treatment equation over gradually expanding subsets of data. We begin by restricting the sample to the pure control group and those participants in the other treatment arms that received their most recent cash transfer or PM+ session 200 or fewer days ago, and then gradually widen this time window by one day and include people whose most recent treatment administration was 201 days ago, then 202 days, and so forth, until we have included the entire sample. The pure control group is always included in its entirety. In each of the nine panels, the horizontal axis corresponds to the time elapsed between endline and the most recent cash transfer or PM+ session, and the vertical axis shows treatment effects estimated for the sample defined by each individual point on the horizontal axis. Note that because the time window is expanding, individual estiamtes are not independent of one another.

G. Demand effects

Figure G.1: Demand effects

	(1) Low demand mean (SD)	(2) High demand mean (SD)	(3) Low demand = high demand (p)	(4) N low demand / N high demand
Number of times beaten or slapped	4.65 (2.31)	4.62 (2.33)	0.89	1545 / 1545
Number of times forced to have intercourse	2.79 (1.25)	2.76 (1.27)	0.68	1545 / 1545
I felt depressed	3.43 (1.08)	3.42 (1.08)	0.71	1545 / 1545

H. Effects of transfer frequency

To assess whether the frequency of cash transfers impacts outcomes, we restrict the sample to participants in CT, CT&PM+, and pure control villages and estimate the following model:

$$y_{vi} = \beta_0 + \beta_1 LumpCT_{vi} + \beta_2 WeeklyCT_{vi} + \beta_3 CT\&PMP_{vi} + \delta y_{viB} + \varepsilon_{vi} \tag{3}$$

Here, $LumpCT_{vi}$ is an indicator that the recipient received the cash transfer all at once; $WeeklyCT_{vi}$ is an indicator the recipient received the cash transfer in weekly installments. PMP_{vi} is an indicator that the respondent received the PM+ program in addition to cash transfers. The comparison between β_1 and β_2 thus identifies the relative effect of lump-sum and installment transfers. The outcomes considered in this specification are the same as for equation 1.[2]

The tables below show OLS estimates of the equation above. Outcome variables, measured at endline, are listed on the left. The first column shows the pure control group mean and standard deviation (in parentheses) of the outcome variable. Columns (2) and (3) show the treatment effects of weekly and lump-sum transfers relative to the pure control group, respectively. Column (4) shows p- values for the comparison of the two treatment effects. Standard errors, clustered at the village level, are shown in parentheses.

* p $<$ 0.1, ** p $<$ 0.05, *** p $<$ 0.01. Standard errors in parentheses.

[2]In the pre-analysis plan, we wrote that this analysis would include village-level fixed effects, thinking that the randomization into lump-sum vs. weekly transfers happened within villages. This was an error—the randomization happened at the village level. We therefore omit the fixed effects, which would be collinear with treatment.

Table H.1: Transfer frequency, primary outcomes

	(1) Pure control mean (SD)	(2) Weekly cash transfers	(3) Lump sum cash transfers	(4) Weekly = lump sum	(5) N
Monthly per-capita non-durable consumption (USD PPP)	52.49	14.93	6.93	0.05**	2397
	(46.51)	(3.90)***	(4.12)*		
		+28%	+13%		
Total value of assets household owns (USD PPP)	553.16	301.67	245.63	0.34	2397
	(664.70)	(52.19)***	(48.41)***		
		+55%	+44%		
Monthly household revenue (USD PPP)	135.48	52.74	20.67	0.02**	2397
	(153.09)	(13.51)***	(13.77)		
		+39%	+15%		
Psychological well-being index	0.00	0.22	0.25	0.71	2397
	(1.00)	(0.07)***	(0.08)***		
Intimate partner violence index	−0.00	−0.00	−0.08	0.50	1394
	(1.00)	(0.10)	(0.09)		

I. Indices and variables

The variables or indices below marked with a $^+$ are primary outcomes of interest. Variables or indices marked with a # are secondary outcomes. Impacts on variables or indices below marked with a * are not primary or secondary, but are reported in the detailed outcome tables.

1. Non-durable consumption$^+$ – monthly consumption per capita
   (a) Food*
      i. Food own production* (value of milk, meat, eggs and additional animal products consumed, value of top three crops produced consumed)
      ii. Food bought
         A. Meat, fish & dairy*
         B. Fruit & vegetables*
         C. Cereals*
         D. Other food* (roots, pulses, oils, sugars, prepared food, other food)
   (b) Temptation good expenditure*
      i. Alcohol
      ii. Tobacco
      iii. Gambling
   (c) Personal and household items*
      i. Clothing and shoes
      ii. Personal items such as soap, shampoo, etc.
      iii. Household items such as matches, kerosene, etc.
   (d) Housing repair or improvement*
   (e) Education expenditures*
   (f) Medical expenditure*
   (g) Social expenditure*
      i. Religious expenses or other ceremonies
      ii. Weddings
      iii. Funerals
   (h) Other expenses greater than KES 1,000
2. Food security# – weighted standardized index of:
   (a) Number of times last month adults cut or skipped meals (negatively coded)*
   (b) Number of times last month children cut or skipped meals (negatively coded)*

(c) Number of times last month had to borrow food or rely on help from a friend or relative (negatively coded)*

(d) All household members eat two meals a day (indicator)*

(e) All household members usually eat until content (indicator)*

(f) Number of times last week respondent has eggs, meat or fish*

3. Assets^{+} – sum of value of:

   (a) Productive assets*

      i. Irrigation pump
      ii. Hose pipe
      iii. Ox-ploughs
      iv. Oxen/work bulls
      v. Knapsack sprayers
      vi. Wheelbarrows
      vii. Ox-carts/donkey carts
      viii. Hand carts
      ix. Other farming tools
      x. Fishing equipment (boats, canoes, etc)
      xi. Other asset used for agriculture or business

   (b) Vehicles*

      i. Bicycle
      ii. Motorbike

   (c) Furniture*

      i. Sofas
      ii. Chairs
      iii. Table
      iv. Clock/Watch
      v. Beds
      vi. Mattresses
      vii. Cupboards
      viii. Other furniture

   (d) Household durables*

i. Cell phone
ii. Sewing machine
iii. Radio, tape or CD player
iv. Battery
v. Solar panel
vi. Television or computer
vii. Kerosene stove
viii. Refrigerator

(e) Other

(f) Livestock*

i. Cows (valued at average sample sales price)
ii. Birds (valued at average sample sales price)
iii. Small ruminants (valued at average sample sales price)

(g) Financial assets* - net balance (KES) of savings minus outstanding loans:

i. Savings with an institution (bank, SACCO, micro-finance organization)
ii. Savings with MPesa
iii. Savings in any other place (e.g., with family or friends)
iv. Loans made by friends or family
v. Loans from moneylenders, micro-finance institutions, shops, banks or other sources

4. Revenue^{+} – sum of monthly household revenue from:

(a) Livestock*

i. Cows

A. Value of milk (sold and consumed)
B. Value of meat (sold and consumed)
C. Value of animals sold
D. Value of other products

ii. Small ruminants

A. Value of meat (sold and consumed)
B. Value of animals sold
C. Value of other products

iii. Birds

A. Value of eggs (sold and consumed)

B. Value of meat (sold and consumed)

C. Value of animals sold

(b) Agricultural income (most recent agricultural season)*

i. Value of crops harvested in most recent season

(c) Enterprise income*

i. Sales in prior month (prorated for share of enterprise owned if applicable)

(d) Wage income*

i. Sum of income from outside labor

5. Profits# – sum of monthly household revenue less costs from:

(a) Livestock revenue - Cost of care (e.g. fodder, veterinary care, etc.) - Cost of rented animals

(b) Crop revenue - Costs of seeds, fertilizers/herbicides/pesticides, hired machines, water, labor and other expenses

(c) Enterprise revenue - Costs of electricity, wages, water, transport, purchase of inputs, other costs (prorated for share of enterprise owned if applicable)

6. Labor# – hours spent per week per capita on income generating activities, including:

(a) Working in agriculture for this household*

(b) Tending animals for this household*

(c) Working in a non-farm or livestock business owned by this household*

(d) Working for pay for someone outside the household (in agriculture, livestock, housework, casual labor, salaried job or other paid work)*

7. Education index# – weighted standardized index of:

(a) Proportion of children (<19) in school*

(b) Average days of school missed per child (<19; last 30 days)* - *negatively coded*

(c) Average spending on school expenses per child (<19; last 12 months)*

(d) Average time studying or in school per child (<19; last 7 days)*

8. Psychological well-being+– weighted standardized index of (a)-(d):

(a) GHQ-12* with standard scoring

(b) Cohen Stress (PSS)*

(c) WVS (happiness)* (1-4 scale)

(d) WVS (life satisfaction)* (1-10 scale)

(e) WHODAS

9. Intimate Partner Violence^{+} – weighted standardized index of physical and sexual violence variables. The recall period is the preceding 6 months.

   (a) Physical violence

      i. How many times per month did your husband beat you, slap you or act violently against you?

      ii. How many times per month did your husband push you, shake you, or throw something at you?

      iii. How many times per month did your husband slap you?

      iv. How many times per month did your husband twist your arm or pull your hair?

      v. How many times per month did your husband punch you with his fist or with something that could hurt you?

      vi. How many times per month did your husband kick you, drag you or beat you up?

      vii. How many times per month did your husband try to choke you or burn you on purpose?

      viii. How many times per month did your husband threaten to attack you with a knife, gun, or any other weapon?

      ix. How many times per month did your husband beat any of the children under the age of 12 living in this household or discipline them using physical measures?

   (b) Sexual violence

      i. How many times per month did your husband physically force you to have sexual intercourse with him even when you did not want to?

      ii. How many times per month did your husband force you to perform any sexual acts you did not want to?

   (c) Emotional violence

      i. How many times per month did your husband not trust you with any money?

ii. How many times per month did your husband expect you to ask him for permission before leaving the house?

iii. How many times per month did your husband ever insult you or make you feel bad about yourself?

(d) Attitudes: Male focused attitudes

i. The important decisions in the family should be made only by the men of the family

ii. The wife has the right to express her opinion even when she disagrees with what her husband is saying

iii. It is the husband's responsibility to provide for the wife and the family

iv. A real man produces a male child

(e) Attitudes about justifiability of violence

i. A wife should tolerate being beaten by her husband

ii. It is acceptable for a husband to not trust his wife with any money

iii. It is acceptable for a husband to expect his wife to ask him for permission before leaving the house

iv. Emotional violence attitude: It is acceptable for a husband to insult his wife or make her feel bad about herself[‡]

v. Physical violence attitude: It is acceptable for a husband to beat his wife, slap her, or act violently against her[‡]

vi. Sexual violence attitude: It is acceptable for a husband to physically force his wife to have sexual intercourse with her even when she does not want to[‡]

vii. Child violence attitude: It is acceptable for a father to beat his children under the age of 12 or discipline them using physical measures.[‡]

(f) Satisfaction with relationship – weighted standardized index of:

i. Are you satisfied with your relationship?

ii. How often do you think your relationship is good compared to most other relationships in your village? How often do you think that your relationship met your original expectations?

iii. How often are there any problems in your relationship?

iv. How often do you wish you hadn't gotten in this relationship?

v. How often do you feel that your husband is unable to provide sufficiently for you and your family?

(g) Violence against children (indicator)

(h) Intimate partner violence: testing for reporting and demand effects

   i. List randomization: physical and sexual violence

   ii. Envelope randomization: physical and sexual violence

   iii. Anonymous reporting by pointing at faces (physical and sexual violence against spouse, physical violence against children)

      A. In the past 6 months, has your husband ever beaten you, slapped you, or acted violently against you? Please point to the happy face if this DID NOT happen to you. Please point to the sad face if this DID happen to you.

      B. In the past 6 months, did your spouse ever beat any of the children under the age of 12 living in this household or disciplined them using physical measures? Please point to the happy face if this DID NOT happen to any children under the age of 12 living in the household. Please point to the sad face if this DID happen to any of the children under the age of 12 living in the household.

      C. In the past 6 months, has your husband ever forced you to have sexual intercourse with him even when you did not want to? Please point to the happy face if this DID NOT happen to you. Please point to the sad face if this DID happen to you.

   iv. Demand treatments: difference in answers to physical and sexual violence questions before vs. after demand treatment, with individual-level fixed effects.

(i) Bargaining power: indifference point between receiving KES 1000 oneself, or one's spouse receiving KES 800–2000.