DISCUSSION PAPER SERIES

DP16640

Mental Health Therapy as a Core Strategy for Increasing Human Capital: Evidence from Ghana

Nathan Barker, Gharad Bryan, Dean Karlan, Angela Ofori-Atta and Christopher Udry

DEVELOPMENT ECONOMICS

LABOUR ECONOMICS

PUBLIC ECONOMICS

CEPR

Mental Health Therapy as a Core Strategy for Increasing Human Capital: Evidence from Ghana

Nathan Barker, Gharad Bryan, Dean Karlan, Angela Ofori-Atta and Christopher Udry

Mental Health Therapy as a Core Strategy for Increasing Human Capital: Evidence from Ghana

Abstract

We study the impact of cognitive behavioral therapy (CBT) for individuals selected from the general population of poor households in rural Ghana. Results from 2-3 months after a randomized intervention show strong impacts on mental and physical health, cognitive and socioemotional skills, and downstream economic outcomes. We find no evidence of heterogeneity by baseline mental distress; we argue that this is because CBT can improve human capital for a general population of poor individuals through two pathways. First, CBT reduces vulnerability to deteriorating mental health; and second, CBT directly improves bandwidth, increasing cognitive and socioemotional skills and hence economic outcomes.

JEL Classification: N/A

Keywords: N/A

Nathan Barker - nathan.barker@yale.edu
Yale University

Gharad Bryan - g.t.bryan@lse.ac.uk
LSE and CEPR

Dean Karlan - dean.karlan@gmail.com
Northwestern University and CEPR

Angela Ofori-Atta - angela.oforiatta@gmail.com
University of Ghana Medical School

Christopher Udry - christopher.udry@northwestern.edu
Northwestern University and CEPR

Mental Health Therapy as a Core Strategy for Increasing Human Capital: Evidence from Ghana

Nathan Barker, Gharad Bryan, Dean Karlan,

Angela Ofori-Atta and Christopher Udry*

October 12, 2021

Abstract

We study the impact of cognitive behavioral therapy (CBT) for individuals selected from the general population of poor households in rural Ghana. Results from 2-3 months after a randomized intervention show strong impacts on mental and physical health, cognitive and socioemotional skills, and downstream economic outcomes. We find no evidence of heterogeneity by baseline mental distress; we argue that this is because CBT can improve human capital for a general population of poor individuals through two pathways. First, CBT reduces vulnerability to deteriorating mental health; and second, CBT directly improves bandwidth, increasing cognitive and socioemotional skills and hence economic outcomes.

Keywords: mental health, poverty, cognitive behavioral therapy, human capital, scarcity
JEL Classification: I15, I31, O12

1   Introduction

Spurred in part by the inclusion of mental health as a key sustainable development goal, a growing "global mental health" movement argues for improved access to therapy in poor countries (e.g., Patel and Prince 2010 and Patel et al. 2018). How broad might the impact of this movement be? We argue that increasing access to mental health therapy in low income countries should be seen as a core means of increasing human capital in the *general* population, with relevance far beyond treating those with a diagnosable mental health condition.

Our argument is based on the results of a large-scale randomized controlled trial evaluating the impact of untargeted, group-based, Cognitive Behavioral Therapy (CBT) in rural Ghana. Using endline data from 2-3 months after the intervention we first show that this program led to meaningful average increases in mental health, socioemotional and cognitive skills, and downstream economic outcomes. For example, those in the treatment group, relative to the control group, report having good mental health 0.53 more days per month; increase self-efficacy by 0.29 standard deviations; improve their score on a digit span test by 0.08 standard deviations; and, report 0.12 standard deviations higher economic status.[1] We then show, perhaps surprisingly, that impacts on mental health, physical health, socioemotional and cognitive skills did not differ based on mental health distress at baseline. These results indicate the program is relevant for a general population, not just those with diagnosed mental health issues. We identify two key mechanisms. First, we argue that low-income individuals are at risk of, or *vulnerable* to, deteriorating mental health, and therapy preemptively alleviates this vulnerability. Second, we argue that CBT has a direct effect on cognitive and socioemotional skills (and then also downstream economic outcomes) even for those who do not or will not suffer from mental health difficulties.

We build our argument for vulnerability in several steps. We first use data from our control group to show that there is substantial churn across mental health states over time. Specifically, 43% of those who report no mental distress at baseline report mental distress at endline 10-12 months later. Meanwhile, about 65% of those that report moderate to severe mental distress at baseline report no mental distress at endline. These figures should be understood in the context of high levels of distress: at baseline 55% have some form of psychological distress and 15% have severe psychological distress. We then show that this churn is *predictable*. We split our control group into training and testing samples, train a Lasso using k-fold cross validation in the training sample, and then predict endline distress in both the control testing set and treatment group. We show that while baseline mental health distress does not predict heterogeneity in

---

[1]Economic status is determined using Cantril's Ladder with respondents asked to assess their current economic wellbeing.

treatment effects, vulnerability to mental distress does. Specifically, mental and physical health impacts of the CBT program are statistically significantly larger for those who are predicted by our Lasso to be vulnerable to deteriorating mental health. This is important as a direct test of our vulnerability hypothesis and also because it indicates that the churn between mental distress states is not just measurement error.

Our argument for a direct effect of CBT, even for those who do not experience mental health challenges, draws on the concept of "bandwidth" defined by Mullainathan and Shafir (2013) and Schilbach, Schofield and Mullainathan (2016). These authors argue that being poor leads people to misallocate their mental resources toward short term financial problems, thus reducing both cognitive and non-cognitive (e.g., self-control) mental capacity available for other tasks. We first review the theory on which CBT is based, and our particular curriculum, and argue that the theoretical mechanism through which CBT is thought to operate suggests that it should engender a better allocation of bandwidth across tasks, drawing a link between therapy and the behavioral economics of scarcity. Second, we show that the CBT program had large impacts on key measures of cognitive and socioemotional skills that are part of bandwidth. Specifically we show a 0.27 standard deviation increase in an index of socioemotional skills including self-control, and a 0.08 standard deviation increase in an index of cognitive skills including classic measures such as digit span and Raven's progressive matrices. Third, we show that while our measure of vulnerability to mental distress predicts the impacts of the CBT on mental and physical health, it *does not* predict impacts on our measures of bandwidth and downstream economic outcomes. This suggests that these less traditional impacts of CBT are created by a different, broader, channel.

Our work builds on several important literatures. Development economists have long recognized that vulnerability is a key part of poverty: being poor not only means having a low income, but also facing frequent negative shocks that threaten to induce a state of destitution (e.g., Morduch 1994, Ligon and Schechter 2003 and Collins et al. 2009). At the same time, a newer literature argues that poverty can lead to mental health difficulties (e.g., Lund et al. 2011, Ridley et al. 2020). Of particular relevance to our argument is the work of Chemin, De Laat and Haushofer (2013) which explicitly shows the negative mental health impact of a *transitory* economic shock. Taken together, the twin claims of vulnerability to economic shocks and a causal effect of shocks on mental health lead to our hypothesis that the poor are vulnerable to mental health difficulties. We contribute to this literature by showing that those vulnerable to mental health deterioration are indeed those who reap the greatest benefits from a CBT program.

Second, a series of papers have argued that poverty changes psychology and decision making beyond mental health. Banerjee and Mullainathan (2010) argues that poverty leads people to give into temptation, Mullainathan and Shafir (2013), Shah et al. (2018) and Schilbach, Schofield and Mullainathan (2016) argue that the poor spend significant

mental resources on short run financial problems reducing bandwidth available for other tasks, and Bessone et al. (2021) argues that the environment in which the poor live directly reduces mental resources. We contribute to this literature by arguing that CBT can be conceptualized as a broad program to improve the quality of decision making, helping individuals to better allocate their mental resources. We also draw a link between this literature and a large literature showing important economic returns to socioemotional, "non-cognitive" skills (Heckman, Stixrud and Urzua 2006, Alan, Boneva and Ertac 2019 and McKelway 2021).

Finally, we contribute to a growing literature that studies the economic impacts of therapy. Existing papers in this literature predominantly study the impact of therapy on a well defined group with a common mental health related difficulty.[2] In contrast, we study the impact of CBT in a general population and argue for a broad relevance of therapy. In this regard, our study is most similar to the contemporaneous work of Haushofer, Mudida and Shapiro (2020). Similar to us, these authors study a CBT program delivered to a general population in a low income country, Kenya. Their results differ markedly from ours. They find no statistically significant impact of CBT on mental health or economic outcomes 12 months after the program. They argue that their program is not successful precisely because it does not target a specific population with a common difficulty to be addressed through therapy. We posit two key differences between their study and ours: intensity of treatment and measurement timeframe. Our program consisted of 12 weekly 90-minute sessions, whereas the Kenya study was 5 weekly 90-minute sessions. Our results show that CBT *can* have impacts on a general population three months after the intervention, while their outcomes are first measured at 12 months, and their null results may be because impacts fade over time. We see this as an important challenge, how can programs be designed to maintain impacts? It could be that repeating or "topping up" therapy may be both effective and low cost, and our results demonstrate an initial first success for such an idea: a first dose of therapy can lead to an initial increase in human capital.

2   The Intervention

2.1   Cognitive Behavioral Therapy

Developed in the 1960s and originally designed to treat depression, CBT is a widely used and widely studied clinical approach to the treatment of multiple mental health conditions. CBT is designed on the premise that individuals have automatic responses to stimuli and that these responses are sometimes subject to "cognitive distortions." These

---

[2]Blattman, Jamison and Sheridan (2017), Bolton et al. (2007), Heller et al. (2017).

distortions in turn lead to the misinterpretation of stimuli, affecting the way people view themselves, others, and the future (Beck 1979). CBT encourages individuals to recognize their automatic responses and question their thought distortions as a way of facilitating productive thinking.

The conceptual framework for CBT gives a clear sense of why the poor might be at both greater risk of mental health difficulties and vulnerable to deteriorating mental health. Those who find themselves in a steady state of poverty are constantly presented with stimuli regarding their own low status, raising significant scope for distortion. For example, an individual born into a poor farming family may misinterpret his low income as evidence of his own low levels of talent leading to significant ongoing mental distress. The poor also face many idiosyncratic shocks in their lives, and there is significant scope for distortions to lead to misinterpretation of the cause of these difficulties. An individual who experiences a bad harvest due to insufficient rainfall might conclude that "my efforts never pay off," or someone who experiences a negative health shock might conclude "no matter what I do, these things always happen to me."[3] These observations are at the core of our claim that CBT may be appropriate for many of the world's poor: large numbers of the world's poor are likely to suffer from poor mental health, and even those who are not currently suffering are likely to fall into poor mental health.

The CBT framework also provides an alternative way to conceptualize the mechanisms that Mullainathan and Shafir (2013) conjecture drive the negative effects of scarcity and trap the poor in poverty. Key to their claim that scarcity leads to negative outcomes must be a notion that responses to scarcity, e.g. "tunneling," or rumination on short-term needs,[4] are a misallocation of mental resources. If this were not the case, these behaviors and the ongoing poverty they cause would be optimal. One way to understand this scarcity-induced misallocation is as an automatic, distorted, response to financial stress. This observation opens the door to think of CBT's focus on automatic thoughts, and explicitly evaluating their accuracy, as a way to learn to avoid the negative outcomes of scarcity induced behaviors, and in particular the resulting decrease in bandwidth. Indeed, several of the key lessons of the CBT curriculum that we use, and CBT in general, address bandwidth reducing behaviors. The manual for the program we study, for example, devotes time to discussing the dangers of: "mental filtering" or dwelling on specific issues; "catastrophising" or over-emphasizing small problems; and "should statements" which require an individual to reach the correct outcome for all problems thus suggesting corner solutions to effort allocation. Thus, while CBT has traditionally been offered to individuals at risk for a specific disorder (Butler et al. 2006), it might also be useful in guiding the automatic response of individuals exposed to stressors on

---

[3]De Quidt and Haushofer (2016) argues for a negative impact of poverty and shocks on mental health even in the absence of the thought distortions that are a mainstay of the therapy literature.

[4]These can be grouped together as a scarcity mindset.

a regular basis, regardless of their current mental health status.

The potential of CBT as a general method to increase human capital is further aided by the fact that group-based CBT is often delivered using a strictly-controlled manual, allowing for CBT to be moved out of a clinical setting. Recent research has demonstrated the ability of lay counselors to deliver CBT to individuals in several low-income countries when targeted at groups with existing mental disorders, such as depressive and anxiety disorders (Patel et al. 2010), perinatal depression (Rahman et al. 2008), or post-traumatic stress disorder (Smith et al. 2007).

2.2 Counselor Characteristics and Training

We study a CBT curriculum designed by one of the authors (Ofori-Atta) and intended to be implemented by recent college graduates with a degree in psychology or a related field and requiring no further qualifications nor training. The study explicitly aims to test whether the program could be integrated with Ghana's National Service Scheme (NSS). The NSS mandates recent college graduates to work for one year in a public service role and is one mechanism through which mental health services are currently delivered in Ghana. In conjunction with Psych Corps Ghana (a program run through The University of Ghana Medical School), recent college graduates with backgrounds in clinical psychology are posted to district hospitals throughout the country (Ofori-Atta, Ketor and Bradley 2014).[5]

The research nonprofit organization Innovations for Poverty Action (IPA) recruited 37 staff to deliver the program, half served as lead counselors, the other half as assistant counsellors. All staff had at least a bachelor's degree (one had an advanced degree), their most common majors were psychology (65%), another health related-field (e.g. community health, 13%), and development studies or social work (13%). The median counsellor member received their tertiary degree two years (mean 2.76 years) prior to being hired.

All counsellors received two weeks of classroom training, and performed one week of piloting. Additionally, at the end of each week, all counselors in a given district met with a lead counselor, who debriefed them on the previous week's activities and helped them prepare for the coming week.

2.3 Curriculum and Program Delivery

The CBT program consisted of 12 weekly 90-minute sessions, delivered to a group of 10, and took place in the community where people lived. The 12 sessions covered four mod-

---

[5]We explored recruiting and staffing counselors through NSS; uncertainty on the timing of our study precluded that approach.

ules: (1) Healthy thinking, including identifying and challenging thought distortions, (2) Solving problems at home and at work (using the skills developed in module 1), (3) Managing relationships, including communication, self-esteem, being good to yourself and others, and (4) Goal-setting and goal-directed behavior. Sessions included a combination of the counselors and assistant counselors introducing the material, having individuals discuss hypothetical scenarios as a group and in pairs, and thinking about how they could apply the lessons they learned to their own lives. As with most CBT interventions, counsellors assigned homework tasks after each session and reviewed these in the next session. The full CBT Manual is available on the authors' websites.

3 Sampling, Randomization, and Data Collection

3.1 Sample Selection and Randomization

The CBT program was part of a larger study that will compare the impact of several interventions, including a cash transfer and a Heifer International graduation program similar to those reported in Banerjee et al. (2015). The data here were collected *prior* to the announcement and implementation of other interventions and hence the impacts of the CBT can be evaluated in isolation.

Appendix A provides details on the sample selection, community criteria, and randomization procedures.

In total, 258 communities satisfied all criteria and were randomized into one of three groups: (1) pure control communities (97 communities), (2) pure CBT communities, in which all households would either be a control household, or only receive CBT (20 communities), or (3) full program communities, in which some households would be in the control group, some would receive CBT, and some would receive a further economic program, delivered after the completion of all activities and surveying discussed in this paper (141 communities).

Within the set of communities chosen to receive CBT only or the full program, half were randomly selected to have their CBT program targeted only at men, and half to have the program targeted only at women. Within the CBT only communities half of the eligible households were randomly allocated to receive CBT and half to receive nothing. Within the full program communities, 11% were chosen to receive CBT and no economic program, 15% CBT and the economic program, 58% the economic program only, and 17% to receive nothing.

3.2 Outcome Data and Indices

The baseline survey consisted of a household and an "adult" survey. The household survey measured consumption, assets and wealth, income, and other household characteristics. The adult survey, administered to the household head and their spouse, measured personal outcomes.[6] At endline we administered only the "adult" survey, again to both the household head and their spouse, roughly two months after the completion of the CBT program (and seven months after the baseline). In our analysis of outcomes we include the responses of both adults in control households; in households where an individual received CBT, we only include treated individuals.[7]

We use the adult survey to create our outcome measures. The household survey variables were used in our re-randomization procedure, and are used to create our measure of vulnerability to deteriorating mental health. We report both outcomes of individual measure and indices that summarize outcomes in five broad categories: mental health; physical health; socioemotional skills; cognitive skill; and economic outcomes. Indices were created as per Kling, Liebman and Katz (2007).

Our mental health index is created from three measures, the Kessler psychological Distress K10 Scale (Kessler et al. 2002), a self rating of mental health taken from the Behavioral Risk Factor Surveillance Survey (BRFSS),[8] and a self report of days in the month without poor mental health. We use the K10 from our baseline survey as the main measure of baseline mental health. Our health index is created from the BRFSS self rating of physical health and a self report of the number of days without poor physical health.[9] It is important to note that improvements of mental health may lead to perceived changes in physical health and hence improvements in self reported physical health.

Our index of non-cognitive or socioemotional skills has three sub-indices: (i) generalized self-efficacy: a measure of optimistic self-belief (Schwarzer, Jerusalem and others 1995); (ii) grit: a measure of passion for and perseverance with long-term goals (Duckworth and Quinn 2009); and (iii) self reported self-control (Tangney, Baumeister and Boone 2004). Four measures comprise our index of cognitive skills: (i) performance on Raven's Progressive Matrices (Raven 1941); (ii) a forward digit span test; (iii) a backwards digit span test; and (iv) a Stroop-like test of executive function (adjusted here for a population with limited literacy; Stroop 1935).

---

[6]In the case of polygynous households one randomly chosen wife was selected to both be surveyed and receive CBT.

[7]I.e. we exclude spouses of individuals who received CBT from our analysis, rather than code them as "treated" or "control."

[8]The question is "In general, would you say your mental health is: excellent, very good, good, fair or poor?"

[9]The BRFSS question is "In general, how would you rate your health?"

Finally, our downstream economic outcomes index is composed of three measures, the number of work days missed due to poor mental or physical health, self-reported economic status reported using Cantril's ladder (Kilpatrick and Cantril 1960), and a self-evaluation of expected economic status in five years (again using Cantril's ladder).

4 Results

4.1 Prevalence and Transition Rates of Psychological Distress

We first show that the poor are vulnerable to psychological distress. Table 1 reports the incidence of psychological distress (measured by the K10 Scale) and transition probabilities into and out of states of psychological distress over the span of seven months in our study sample (Panels A and B), and over four years in a similar population from the Ghana Socioeconomic Panel Survey (Panel C). Despite not sampling based on existing mental health, the rate of psychological distress is high, with 55% reporting symptoms associated with some degree of psychological distress (compared to 57% in the general population, Panel C). To compare, in the United States, the 2007 Behavioral Risk Factor Surveillance Survey (BRFSS) documents only 13% with any level of psychological distress (Dhingra et al. 2011).

Our assertion that CBT is applicable as a mental health intervention for individuals not currently experiencing mental illness depends in part on the observation that low-income individuals diagnosed as "well" at a given point in time are nonetheless at elevated risk for subsequent transitions into psychological distress. The high degree of churn into and out of psychological distress shown in Table 1 supports this view. Among individuals observed to have no psychological distress at baseline, 43% have some form of distress at endline; 10% have severe psychological distress. In fact, of the 16.2% of individuals whose symptoms suggest severe psychological distress at endline, a roughly equal number come from individuals whose baseline responses indicate no distress as those with responses indicating severe psychological distress (0.45 well at baseline * 0.10= 0.045; 0.15 with severe psychological distress at baseline * 0.27=0.041). Our results suggest a mental health program restricted to individuals with existing psychological distress may miss a large number of at-risk, or vulnerable, individuals.

One might be concerned that these transitions merely represented measurement error. Two considerations mitigate this concern. First, if eligibility for CBT were determined by mental distress, and the measurement error is classical, then there would be a large degree of mis-targeting, reiterating our point that the general population would benefit. Second, we will show below that CBT treatment effects are heterogeneous in predicted changes in mental health, indicating that not all transitions can be attributed to noisy measurement of a stable characteristic.

4.2 Average Treatment Effects and Heterogeneity by Baseline Mental Distress

Our main impact estimates are based on comparing those assigned to receive CBT to those not (i.e., pooling the control group households in CBT treatment communities with the control group households in control communities).

Our main results, reported in Tables 2 and 3,[10] show impacts of CBT on mental and physical health, cognitive and socioemotional skills ("bandwidth") and proximate economic outcomes. We estimate average treatment effects in column (2) using the specification

$$y_{ivt} = \alpha + \beta_1 \cdot CBT_{ivt} + \beta_2 \cdot y_{iv0} + X_{ivt}\Pi + \theta_v + \epsilon_{ivt},$$

where y_{ivt} is an outcome variable for individual i in village v at time period t, CBT_{ivt} is an indicator variable for being offered the CBT program, y_{iv0} is the outcome of interest at baseline,[11] X_{ivt} are the variables used in the re-randomization procedure (listed in Appendix Table 1), and θ_v are village dummies.

Columns (3) to (5) present heterogeneous treatment effects and tests for equality by baseline psychological distress.[12] Columns (6) to (8) present heterogeneity by gender.[13] These estimates come from a regression of the form

$$y_{ivt} = \alpha + \beta_1 \cdot CBT_{ivt} \cdot 1_{iv} + \beta_2 \cdot CBT_{ivt} \cdot 2_{iv} + \beta_3 \cdot 1_{iv} + \beta_4 \cdot y_{iv0} + X_{ivt}\Pi + \theta_v + \epsilon_{ivt},$$

where J_{ivt}; $J \in \{1,2\}$ indicates the margin on which we are testing for heterogeneity (i.e. baseline distress, gender).

Given the multi-stage nature of our randomization, we use randomization inference to test our null hypotheses of (i) no treatment effect and (ii) no heterogeneity of treatment effects by baseline distress and gender.[14] Specifically, for each of the randomizations we initially performed to determine an individual's final treatment status (community-level randomization, gender of CBT recipients in a community, individuals assignment) we replicate our initial procedure, and using the same re-randomization selection process, assign placebo treatments. Following this placebo assignment, we test for average

---

[10]Appendix Tables 2-6 show that results are similar if we look only at the difference in outcomes between CBT households and control households in the pure control communities, i.e. excluding from the analysis the control households in communities where some households received CBT; results are robust to this, as evidence of spillovers are small and generally not statistically significant.

[11]When baseline measures are missing they are coded as 0 with an indicator variable for "missing baseline value"

[12]We test for heterogeneity using a binary indicator of any psychological distress to maximize our statistical power to detect such an effect. Our results do not differ (i.e. we do not see evidence of heterogeneity) when using a different threshold, nor when using our continuous measure of baseline Kessler Score).

[13]Reported since the program was delivered to groups of a single gender in each community.

[14]Results are extremely similar when we instead cluster at the village level; available upon request

treatment effects and heterogeneity of the (placebo) treatment by gender and baseline distress. We perform this procedure 2,000 times (following the best practice laid out by Young 2019) and compare the distribution of coefficients (and differences in coefficients for measures of heterogeneity) from these placebo assignments to our coefficients from our true treatment status. Our "RI p-values" report the results of this procedure.[15]

Table 2 reports effects of the CBT intervention on mental and physical health outcomes. We find that CBT led to large improvements in mental and physical health. We see a statistically significant 0.15 standard deviation improvement in our mental health summary index. Breaking this down, individuals receiving CBT have lower scores on the Kessler K10 Scale, are 10% (6pp, p-val=0.004) less likely to have any psychological distress, 20% (6pp, p-val=0.000) less likely to have moderate psychological distress, and 24% (4pp, p-val=0.008) less likely to have severe psychological distress. Individuals also report an 11% reduction in the number of days with poor mental health (0.53 days, p-val=0.097). We see a statistically significant 0.15 standard deviation improvement in our index of physical health. Individuals who received CBT also rated themselves higher on the BRFSS self-report questions on mental and physical health. While these measures are self-reports, previous studies have suggested that these measures meaningfully correlate with real-world health outcomes.[16] Finally, there is a 20% reduction in the number of days with poor physical health (0.89 days, p-val=0.000).

We see limited evidence of heterogeneity by baseline psychological distress. For each of the outcomes reported in Table 2 we are not able to reject equality of treatment effects at the 10% level (column 5); in two cases the estimates approach statistical significance (p-vals = 0.11, 0.17), but even in these two cases the treatment effect is *larger* among individuals scored as "well" at baseline. We are also consistently able to reject the null that there are no impacts of CBT on mental health outcomes for both sub-groups (all 10 outcomes for those well at baseline, 8 of 10 for those with distress at baseline). This is consistent with the idea that a proportion of those who were not distressed went on to become distressed and hence CBT was valuable for them, and also that a proportion of those that were distressed were going to recover even in the absence of the intervention.

We are also uniformly unable to reject the hypothesis that treatment effects are the same for men and women. Moreover, for both genders, we are able to reject the null of no treatment effects for both the mental health and physical health indices, suggesting the effects are not concentrated among either gender.

Table 3 tests our hypothesis that CBT can improve the socio-emotional skills and bandwidth of low-income individuals, and that in turn, improvements in mental health

---

[15]This procedure is described in greater depth in Appendix C.

[16]See for example, Case and Deaton (2020), or Idler and Benyamini (1997), the latter of whom documents that health self-report questions predict mortality in twenty-seven countries, even after controlling for various objective measures of health.

and bandwidth can shape economic behavior. Panel A shows that the treatment led to a 0.27 standard deviation improvement in our index of socio-emotional skills. The CBT program led to improvements in all three of our sub-measures: generalized self-efficacy; grit; and self-control. In Panel B, we see a more modest but still statistically significant 0.08 standard deviation increase in our cognition index. This smaller effect is consistent with the perceived wisdom that cognitive skills are harder to move in a sample of adults. We observe statistically significant positive treatment effects on two sub-measures of cognitive performance: the forward and backwards digit span tests. We are unable to reject the null of no impact on Raven's Progressive Matrices or a Stroop test. Once again, we do not see evidence of heterogeneous treatment effects by baseline psychological distress; CBT led to improvements on these measures both for individuals with and without distress at baseline. We also see little evidence of heterogeneity by gender for these outcomes.

Panel C shows proximate economic outcomes of the CBT intervention. We see a statistically significant 0.19 standard deviation improvement in our index of economic outcomes. Breaking this down, two mechanisms through which depression has been hypothesized to affect economic productivity are through increasing the psychic cost of effort, and through distorted (negative) thoughts about the future. We find evidence of improvements in the second domain, but are unable to reject the null of no improvements in labor supply as a result of the program. In particular, individuals report expecting to be 0.35 (p-val=0.000) points higher on a ten-point Cantril's economic ladder in five years time. On average, individuals report 0.37 fewer days in which poor mental or physical health kept them from engaging in their regular activities, including work and self-care, but this result is not statistically significant in our randomization inference procedure (p-val=0.140).[17] There is some evidence here that impacts are concentrated among the sub-sample with psychological distress at baseline. For example, on our measure of days in which poor health kept individuals from engaging in their regular activities, we observe a treatment effect of 0.48 days (p-val=0.097) for those with distress, and 0.01 (p-val=0.98) for those without, although this difference is not statistically significant at conventional levels. For none of these outcomes are we individually able to reject the null of equal treatment effects, but we do see that our summary index's treatment effects are concentrated among individuals with psychological distress at baseline. Again, there is little to suggest heterogeneity by gender.

---

[17]When estimating using standard errors clustered at the village level, our p-value is 0.10; this discrepancy is consistent with this reuslt being relatively "high leverage" (many 0s, some 30s).

4.3 Heterogeneity by Predicted Endline Psychological Distress

Next we present a more direct test of our vulnerability hypothesis by evaluating whether individuals predicted to fall into distress benefit more from CBT. Specifically, we use a least absolute shrinkage and selection operator (LASSO) with a holdout sample to assign to each individual a "psychological distress risk score", and test for heterogeneous treatment effects with respect to the score. For each of 1,000 repetitions, we perform the following steps:

1. Randomly split the *control* group in half, into a testing and training sample;

2. Estimate a distress risk score on the training sample in the control group, using k-fold cross validation LASSO;[18]

3. Using the results of the LASSO, estimate a fitted-value distress risk score for all individuals in the treatment group, and control individuals in the testing group;

4. Using all treatment individuals and the testing sample of control individuals, regress outcomes of interest on the distress risk score, CBT Treatment, and the interaction of treatment and the distress risk score

$$y_{ivt} = \alpha + \beta_1 \cdot CBT_{ivt} + \beta_2 \cdot \widehat{risk\ score}_{ivt} + \beta_3 \cdot \widehat{risk\ score}_{ivt} \cdot CBT_{ivt} + \epsilon_{ivt}.$$

Table 4, Panel A, presents results with our five summary indices as left hand side variables, as well as the Kessler Score and days in which poor physical and mental health affected individuals' ability to do their regular activities. Coefficients are the median coefficient across the 1,000 repetitions.[19] Following Chernozhukov et al. (2018), we report confidence intervals for the range $1 - 2\alpha$ (rather than $1 - \alpha$) to reflect sampling uncertainty as a result of splitting the control group into a training and testing sample.

CBT is more effective at reducing psychological distress and improving mental and physical health outcomes for those with a high predicted distress risk score. For every 10 point increase in the predicted endline Kessler score the program leads to a 2.7 point reduction in the K10 score, and about a 0.3 std deviation improvement in mental and physical health. These are sizeable effects, for example for the observed Kessler Score, our regression coefficients imply an estimated treatment effect of 0.09 standard deviations (0.71 Kessler points) for someone at the 10th percentile of predicted endline Kessler score, as compared to an estimated treatment effect of 0.28 standard deviations (2.18 Kessler points) for someone at the 90th percentile of predicted Kessler Score.[20]

---

[18]In particular, we use the user-written Stata package lassopack (Ahrens, Hansen and Schaffer 2019)

[19]The share of repetitions in which each variable is selected by the LASSO is reported in Appendix Table 7.

[20]90/10 comparisons for all the variables can be found in the final two rows Table 4.

In contrast, we find no statistically significant heterogeneity for our indices of economic outcomes, socioemotional skills and cognition. 90/10 differences for these indices are uniformly small, and for two of the indices go in the wrong direction. Finally, in Column 7 we report on heterogeneity in days in which poor mental and physical health restrict activity. This outcome is directly downstream of mental health, and not predicted to change because of increased bandwidth. Here we see some evidence of heterogeneity: someone at the 10th percentile of predicted endline Kessler score sees a 0.07 decreased in missed days, while for someone at the 90th percentile this becomes a 0.73 day decrease. Despite this large heterogeneity in means, the coefficient on the interaction between treatment and predicted endline Kessler scores is not statistically significant.

How should the null effects for downstream bandwidth and economic outcomes be interpreted? A growing literature has argued that poor mental health can cause a worsening of economic outcomes. If this hypothesis were true, then we would expect that the strong heterogeneous treatment effects that we see for mental health would follow through to our downstream outcomes. That we do not see that heterogeneity suggests that there is an alternative causal path between CBT and economic, cognitive and socio-emotional outcomes. Overall these results are consistent with our twin hypotheses: CBT has broad applicability in low income communities because many people are likely to fall into poor mental health, and CBT has impacts on human capital directly through an improvement in cognitive and socioemotional skills.

We have not hypothesized that bandwidth is a fast moving variable similar to mental health. Our reading of the literature is that it is the constant grind of poverty that is hypothesized to lead to low bandwidth, rather than short run shocks. We see the direct impacts of CBT on bandwidth variables as consistent with this view, and the lack of heterogeneity with respect to predicted distress as showing that these bandwidth variables are directly impacted by CBT. Nevertheless, for completeness we report a similar analysis where we use a LASSO to create a risk measure for our bandwidth outcomes. We report more fully on this exercise in Appendix Section B and provide results in Appendix Table 8. Consistent with our theory, we do not see heterogeneous treatment effects by predicted change in bandwidth.

5   Conclusion

We find that a Cognitive Behavioral Therapy program, delivered by non-specialist providers in a low-income population in Ghana, reduces psychological distress, improves self-reported mental and physical health, increases bandwidth, and improves short-term economic outcomes. We argued that the results, albeit measured at a short-time horizon of two to three months post intervention, are suggestive of a bipartite expansion

of the domain of applicability for CBT: the poor are vulnerable to mental health problems and CBT can successfully inoculate a broad proportion of the population against the possibility of future mental health problems; and the poor can generally benefit from CBT whether they have mental health problems or not, because CBT improves socio-emotional skills and bandwidth.

Our results also corroborate previous work (e.g. Singla et al. 2017) showing that therapy can be delivered successfully by non-specialist providers in low-income countries. We show this pattern holds in a large sample when delivered to a general low-income population, rather than targeted at a specific form of mental illness.

Future research should determine the extent to which impacts manifest themselves in the long-run, and if impacts fade, should investigate strategies to improve long term impact.

References

Ahrens, Achim, Christian B Hansen, and Mark Schaffer. 2019. "LASSOPACK: Stata module for lasso, square-root lasso, elastic net, ridge, adaptive lasso estimation and cross-validation."

Alan, Sule, Teodora Boneva, and Seda Ertac. 2019. "Ever failed, try again, succeed better: Results from a randomized educational intervention on grit." *The Quarterly Journal of Economics*, 134(3): 1121–1162.

Asiedu, Edward, Dean Karlan, Monica Lambon-Quayefio, and Christopher Udry. 2020. "A call for structured ethics appendices in social science papers." *Proceedings of the National Academy of Sciences*, 118(29).

Banerjee, Abhijit, and Sendhil Mullainathan. 2010. "The shape of temptation: Implications for the economic lives of the poor." National Bureau of Economic Research Working Paper 15973.

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).

Beck, Aaron T. 1979. *Cognitive therapy of depression.* Guilford press.

Bessone, Pedro, Gautam Rao, Frank Schilbach, Heather Schofield, and Mattie Toma. 2021. "The economic consequences of increasing sleep among the urban poor." *The Quarterly Journal of Economics*, 136(3): 1887–1941.

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.

Bolton, Paul, Judith Bass, Theresa Betancourt, Liesbeth Speelman, Grace Onyango, Kathleen F Clougherty, Richard Neugebauer, Laura Murray, and Helen Verdeli. 2007. "Interventions for depression symptoms among adolescent survivors of war and displacement in northern Uganda: a randomized controlled trial." *JAMA*, 298(5): 519–527.

Butler, Andrew C, Jason E Chapman, Evan M Forman, and Aaron T Beck. 2006. "The empirical status of cognitive-behavioral therapy: a review of meta-analyses." *Clinical Psychology Review*, 26(1): 17–31.

Case, Anne, and Angus Deaton. 2020. *Deaths of Despair and the Future of Capitalism.* Princeton University Press.

Chemin, Matthieu, Joost De Laat, and Johannes Haushofer. 2013. "Negative rainfall shocks increase levels of the stress hormone cortisol among poor farmers in Kenya." *SSRN*, 2294171.

Chernozhukov, Victor, Mert Demirer, Esther Duflo, and Ivan Fernandez-Val. 2018. "Generic machine learning inference on heterogenous treatment effects in randomized experiments." National Bureau of Economic Research Working Paper 24678.

Collins, Daryl, Jonathan Morduch, Stuart Rutherford, and Orlanda Ruthven. 2009. *Portfolios of the poor: how the world's poor live on $2 a day.* Princeton University Press.

De Quidt, Jonathan, and Johannes Haushofer. 2016. "Depression for economists." National Bureau of Economic Research Working Paper 22973.

Dhingra, Satvinder S, Matthew M Zack, Tara W Strine, Benjamin G Druss, Joyce T Berry, and Lina S Balluz. 2011. "Psychological distress severity of adults reporting receipt of treatment for mental health problems in the BRFSS." *Psychiatric Services*, 62(4): 396–403.

Duckworth, Angela Lee, and Patrick D Quinn. 2009. "Development and validation of the Short Grit Scale (GRIT–S)." *Journal of Personality Assessment*, 91(2): 166–174.

Haushofer, Johannes, Robert Mudida, and Jeremy P Shapiro. 2020. "The Comparative Impact of Cash Transfers and a Psychotherapy Program on Psychological and Economic Well-being." National Bureau of Economic Research Working Paper 28106.

Heckman, James J, Jora Stixrud, and Sergio Urzua. 2006. "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior." *Journal of Labor Economics*, 24(3): 411–482.

Heller, Sara B, Anuj K Shah, Jonathan Guryan, Jens Ludwig, Sendhil Mullainathan, and Harold A Pollack. 2017. "Thinking, fast and slow? Some field experiments to reduce crime and dropout in Chicago." *The Quarterly Journal of Economics*, 132(1): 1–54.

Idler, Ellen L, and Yael Benyamini. 1997. "Self-rated health and mortality: a review of twenty-seven community studies." *Journal of health and social behavior*, 21–37.

Kessler, R C, Gavin Andrews, L J Colpe, Eva Hiripi, D K Mroczek, S L T Normand, E E Walters, and A M Zaslavsky. 2002. "Short screening scales to monitor population prevalences and trends in non-specific psychological distress." *Psychological Medicine*, 32(6): 959.

Kilpatrick, Franklin Pierce, and Hadley Cantril. 1960. "Self-anchoring scaling: A measure of individuals' unique reality worlds." *Journal of Individual Psychology*, 16(2): 158.

Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz. 2007. "Experimental analysis of neighborhood effects." *Econometrica*, 75(1): 83–119.

Ligon, Ethan, and Laura Schechter. 2003. "Measuring vulnerability." *The Economic Journal*, 113(486): C95–C102.

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.

MacKay, Douglas. 2018. "The ethics of public policy RCTs: The principle of policy equipoise." *Bioethics*, 32(1): 59–67.

McKelway, Madeline. 2021. "Women's Agency and Women's Employment: How Women's Sense of Agency Affects Their Labor Supply." *Unpublished manuscript*.

Morduch, Jonathan. 1994. "Poverty and vulnerability." *The American Economic Review*, 84(2): 221–225.

Mullainathan, Sendhil, and Eldar Shafir. 2013. *Scarcity: Why having too little means so much.* Macmillan.

Ofori-Atta, A, R Ketor, and E Bradley. 2014. "Positioning a new cadre of community workers in the mental health system of a low-resource country: The case of Ghana." *South African Journal of Psychiatry*, 20(3): 105–106.

Patel, Vikram, and Martin Prince. 2010. "Global mental health: a new global health field comes of age." *JAMA*, 303(19): 1976–1977.

Patel, Vikram, Helen A Weiss, Neerja Chowdhary, Smita Naik, Sulochana Pednekar, Sudipto Chatterjee, Mary J De Silva, Bhargav Bhat, Ricardo Araya, Michael King, and others. 2010. "Effectiveness of an intervention led by lay health counsellors for depressive and anxiety disorders in primary care in Goa, India (MANAS): a cluster randomised controlled trial." *The Lancet*, 376(9758): 2086–2095.

Patel, Vikram, Shekhar Saxena, Crick Lund, Graham Thornicroft, Florence Baingana, Paul Bolton, Dan Chisholm, Pamela Y Collins, Janice L Cooper, Julian Eaton, et al. 2018. "The Lancet Commission on global mental health and sustainable development." *The Lancet*, 392(10157): 1553–1598.

Rahman, Atif, Abid Malik, Siham Sikander, Christopher Roberts, and Francis Creed. 2008. "Cognitive behaviour therapy-based intervention by community health workers for mothers with depression and their infants in rural Pakistan: a cluster-randomised controlled trial." *The Lancet*, 372(9642).

Raven, John C. 1941. "Standardization of progressive matrices, 1938." *British Journal of Medical Psychology*, 19(1): 137–150.

Ridley, Matthew, Gautam Rao, Frank Schilbach, and Vikram Patel. 2020. "Poverty, depression, and anxiety: Causal evidence and mechanisms." *Science*, 370(6522).

Schilbach, Frank, Heather Schofield, and Sendhil Mullainathan. 2016. "The psychological lives of the poor." *American Economic Review*, 106(5): 435–40.

Schwarzer, Ralf, Matthias Jerusalem, and others. 1995. "Generalized self-efficacy scale." *Measures in Health Psychology: A User's portfolio. Causal and Control Beliefs*, 1(1): 35–37.

Shah, Anuj K, Jiaying Zhao, Sendhil Mullainathan, and Eldar Shafir. 2018. "Money in the mental lives of the poor." *Social Cognition*, 36(1): 4–19.

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: 149–181.

Smith, Patrick, William Yule, Sean Perrin, Troy Tranah, TIM Dalgleish, and David M Clark. 2007. "Cognitive-behavioral therapy for PTSD in children and adolescents: a preliminary randomized controlled trial." *Journal of the American Academy of Child & Adolescent Psychiatry*, 46(8): 1051–1061.

Stroop, J Ridley. 1935. "Studies of interference in serial verbal reactions." *Journal of Experimental Psychology*, 18(6): 643.

Tangney, June P, Roy F Baumeister, and Angie Luzio Boone. 2004. "High self-control predicts good adjustment, less pathology, better grades, and interpersonal success." *Journal of Personality*, 72(2): 271–324.

Young, Alwyn. 2019. "Channeling fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results." *The Quarterly Journal of Economics*, 134(2): 557–598.

Tables

Table 1: Incidence and Transition Rates of Mental Distress

Panel A: Transition Matrix for Control Group

Level of Baseline Mental Distress, Control Group		Endline Mental Distress				
	(1)	(2)	(3)	(4)	(5)	(6)
	Share at baseline	No Mental distress	Mild Mental Distress	Moderate Mental Distress	Severe Mental Distress	Total
(a) No baseline mental distress	0.45	0.57	0.19	0.14	0.10	100%
(b) Mild baseline mental distress	0.24	0.42	0.24	0.17	0.17	100%
(c) Moderate baseline mental distress	0.16	0.35	0.24	0.18	0.22	100%
(d) Severe baseline mental distress	0.15	0.31	0.23	0.20	0.27	100%
(e) Share at endline		0.46	0.22	0.16	0.16	

Share above diagonal (worsened mental health)	0.31	
Share at diagonal (no change in mental health)	0.38	
Share below diagonal (improved mental health)	0.31	

Panel B: Treatment Effects for each transition cell

		CBT Treatment Effects, by Cell				
	(1)	(2)	(3)	(4)	(5)	(6)
	Control share at baseline	No Mental distress	Mild Mental Distress	Moderate Mental Distress	Severe Mental Distress	Total
(a) No baseline mental distress	0.45	0.06	0.01	-0.05	-0.01	0.00
(b) Mild baseline mental distress	0.24	0.05	-0.01	0.01	-0.04	0.00
(c) Moderate baseline mental distress	0.16	0.08	0.02	-0.01	-0.08	0.00
(d) Severe baseline mental distress	0.15	0.05	0.02	0.00	-0.07	0.00
P-value of Test: Share above diagonal equal for both groups	0.001					

Panel C: Means and Transition Probabilities, Ghana Socio-Economic Panel Survey, Northern, Upper East, Brong Ahafo, Ashanti Regions, non-Regional Capitals

Level of 2009 Mental Distress, Control Group		2013 Mental Distress				
	(1)	(2)	(3)	(4)	(5)	(6)
	Share 2009	No Mental distress	Mild Mental Distress	Moderate Mental Distress	Severe Mental Distress	Total
(a) No 2009 mental distress	0.43	0.70	0.19	0.08	0.03	100%
(b) Mild 2009 mental distress	0.30	0.66	0.21	0.10	0.03	100%
(c) Moderate 2009 mental distress	0.17	0.59	0.25	0.12	0.04	100%
(d) Severe 2009 mental distress	0.11	0.61	0.23	0.13	0.04	100%
(e) Share in 2013		0.66	0.21	0.10	0.03	

Share above diagonal (worsened mental health)	0.17	
Share at diagonal (no change in mental health)	0.39	
Share below diagonal (improved mental health)	0.44	

Table 2: CBT Treatment Effects - Health Outcomes

	Average Treatment Effects		Heterogeneity by Baseline Mental Distress			Heterogeneity by gender of recipient		
	Control Mean	CBT Average Treatment Effect, Full Sample	CBT Average Treatment Effect, Minor, Moderate or Severe Baseline Distress	CBT Average Treatment Effect, No Baseline Distress	p-value from Test: Homogenous Treatment Effect by Baseline Distress, 3=4	CBT Average Treatment Effect, Female	CBT Average Treatment Effect, Male	p-value from Test: Homogenous Treatment Effect by Gender, 6=7
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Mental Health Outcomes								
Mental Health Index	0.00	0.15	0.11	0.18		0.13	0.16	
RI p-value		0.000	0.009	0.010	0.373	0.009	0.007	0.784
Kessler Score	21.39	-1.36	-1.09	-1.63		-1.33	-1.36	
RI p-value		0.000	0.001	0.004	0.391	0.004	0.003	0.962
No distress (Kessler < 20)	0.45	0.06	0.05	0.05		0.06	0.05	
RI p-value		0.004	0.029	0.122	0.960	0.029	0.072	0.930
No moderate or severe distress (Kessler < 25)	0.69	0.06	0.05	0.07		0.06	0.06	
RI p-value		0.000	0.012	0.040	0.585	0.023	0.031	0.924
No severe distress (Kessler <30)	0.85	0.04	0.02	0.07		0.04	0.03	
RI p-value		0.008	0.246	0.019	0.111	0.066	0.108	0.840
Mental Health Self Rating (1/4)	2.84	0.07	0.07	0.04		0.03	0.12	
RI p-value		0.047	0.084	0.475	0.691	0.604	0.030	0.252
Days in month without poor health	25.29	0.53	0.22	1.19		0.77	0.29	
RI p-value		0.097	0.536	0.056	0.169	0.083	0.522	0.471
Panel B: Physical Health Outcomes								
Physical Health Index	0.00	0.15	0.13	0.19		0.15	0.16	
RI p-value		0.000	0.001	0.004	0.447	0.002	0.001	0.854
Physical Health Self-Rating (1/4)	3.03	0.12	0.10	0.14		0.10	0.15	
RI p-value		0.000	0.002	0.009	0.520	0.015	0.000	0.405
Days in Month without poor physical health	25.57	0.89	0.70	1.14		1.03	0.74	
RI p-value		0.000	0.038	0.055	0.549	0.004	0.080	0.608

Each cell in Column 2 is from a single specification estimating the Intent to Treat treatment effect, with between 7,412 and 7,445 observations. Each row for Columns 3-4 are from a single specification with between 6,899 and 6,927 observations, which include a dummy variables for baseline psychological distress and interactions of being randomized into the CBT program interacted with whether the individual had psychologoical distress at baseline. Column 5 reports the p-value from the test that the coefficients in columns 3 and 4 are equal. Each row for Columns 6-7 are from a single specification with between 6,899 and 6,927 observations, which include dummy variables for gender and interactions of being randomized into the CBT program interacted with the gender of the recipient. Column 8 reports the p-value from the test that the coefficients in columns 6 and 7 are equal. P-values are all calculated via randomization inference; the full procedure is described in Appendix A.

Table 3: CBT Treatment Effects - Bandwidth and Economic Outcomes

	Average Treatment Effects		Heterogeneity by Baseline Mental Distress			Heterogeneity by gender of recipient		
	Control Mean	CBT Average Treatment Effect, Full Sample	CBT Average Treatment Effect, Minor, Moderate or Severe Baseline Distress	CBT Average Treatment Effect, No Baseline Distress	p-value from Test: Homogenous Treatment Effect by Baseline Distress, 3=4	CBT Average Treatment Effect, Female	CBT Average Treatment Effect, Male	p-value from Test: Homogenous Treatment Effect by Gender, 6=7
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Socio-Emotional Skills								
Socio-Emotional Skill Index	0.00	0.27	0.25	0.29		0.25	0.28	0.61
RI p-value		0.000	0.000	0.000	0.593	0.000	0.000	0.725
Generalized Self-Efficacy Score	0.00	0.29	0.29	0.30		0.27	0.31	
RI p-value		0.000	0.000	0.000	0.854	0.000	0.000	0.641
Grit Score	0.00	0.19	0.18	0.20		0.20	0.17	
RI p-value		0.000	0.000	0.004	0.835	0.000	0.007	0.794
Self-Control Score	0.00	0.12	0.09	0.16		0.09	0.15	
RI p-value		0.004	0.052	0.024	0.438	0.138	0.022	0.516
Panel B: Cognition								
Cognition Index	0.00	0.08	0.08	0.08		0.04	0.12	
RI p-value		0.012	0.042	0.181	0.996	0.402	0.036	0.304
Raven's Progressive Matrices, Indexed	0.00	0.03	0.02	0.08		0.06	-0.02	
RI p-value		0.537	0.711	0.249	0.475	0.282	0.713	0.352
Digit Span: Forwards, Indexed	0.00	0.08	0.08	0.04		0.03	0.12	
RI p-value		0.025	0.063	0.541	0.581	0.630	0.028	0.232
Digit Span: Backwards, Indexed	0.00	0.07	0.05	0.08		0.03	0.09	
RI p-value		0.036	0.204	0.170	0.697	0.437	0.079	0.405
Executive Function Test, Indexed	0.00	0.05	0.06	0.03		0.00	0.11	
RI p-value		0.153	0.210	0.666	0.737	0.946	0.060	0.156
Panel C: Economic Outcomes								
Economic Index	0.00	0.19	0.20	0.08		0.19	0.19	
RI p-value		0.000	0.000	0.245	0.127	0.001	0.002	0.979
Days in which poor mental or physical health did not keep individual from doing regular activities	26.86	0.37	0.48	0.01		0.49	0.25	
RI p-value		0.140	0.097	0.984	0.428	0.160	0.528	0.632
Self-Reported Economic Status	3.08	0.43	0.44	0.23		0.42	0.44	
RI p-value		0.000	0.000	0.131	0.225	0.002	0.002	0.901
Projected Economic Status in 5 years	5.79	0.35	0.38	0.16		0.30	0.39	
RI p-value		0.000	0.004	0.347	0.312	0.038	0.015	0.700

Each cell in Column 2 is from a single specification estimating the Intent to Treat treatment effect, with between 7,412 and 7,445 observations. Each row for Columns 3-4 are from a single specification with between 6,899 and 6,927 observations, which include a dummy variables for baseline psychological distress and interactions of being randomized into the CBT program interacted with whether the individual had psychologoical distress at baseline. Column 5 reports the p-value from the test that the coefficients in columns 3 and 4 are equal. Each row for Columns 6-7 are from a single specification with between 6,899 and 6,927 observations, which include dummy variables for gender and interactions of being randomized into the CBT program interacted with the gender of the recipient. Column 8 reports the p-value from the test that the coefficients in columns 6 and 7 are equal. P-values are all calculated via randomization inference; the full procedure is described in Appendix A.

Table 4: Heterogeneous Effects by LASSO-Predicted Depression Risk Score

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Kessler Psychological Distress Score	Mental Health Index	Physical Health Index	Economic outcomes Index	Socioemotional Skills Index	Cognition Index	Days in which poor mental or physical health did not keep individual from doing regular activities
Assigned to CBT	0.360	-0.072	-0.068	0.277	0.115	0.159	0.411
	[-0.945, 1.667]	[-0.237, 0.094]	[-0.225, 0.088]	[0.093, 0.458]	[-0.079, 0.307]	[-0.023, 0.349]	[-0.690, 1.527]
Unadjusted p-value	0.606	0.467	0.467	0.013	0.335	0.151	0.529
P-Value adjusted for sample split	1.000	0.934	0.934	0.026	0.670	0.302	1.000
Predicted Kessler Score from Baseline Covariates							
(rescaled to minimum area of common support)	1.066	-0.137	-0.117	-0.072	-0.064	-0.041	0.554
	[0.955, 1.174]	[-0.149, -0.125]	[-0.129, -0.104]	[-0.085, -0.059]	[-0.078, -0.050]	[-0.052, -0.029]	[0.470, 0.637]
Unadjusted p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000
P-Value adjusted for sample split	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Assigned to CBT x Predicted Kessler Score	-0.266	0.034	0.032	-0.012	0.017	-0.012	-0.120
	[-0.452, -0.076]	[0.009, 0.058]	[0.009, 0.055]	[-0.038, 0.014]	[-0.009, 0.044]	[-0.037, 0.013]	[-0.289, 0.049]
Unadjusted p-value	0.021	0.027	0.022	0.437	0.284	0.440	0.246
P-Value adjusted for sample split	0.042	0.054	0.044	0.874	0.568	0.88	0.492
Estimated treatment effect: individual with 10th percentile predicted Kessler Score	-0.712	0.065	0.063	0.225	0.186	0.112	-0.071
Estimated treatment effect: individual with 90th percentile predicted Kessler Score	-2.179	0.250	0.242	0.158	0.281	0.047	-0.730

In each simulation, the control group is divided in two groups: a training control sample and holdout control. The LASSO is estimated using k-fold estimation on the training control, predicting endline Kessler score using baseline covariates. The above specification to estimate treatment effect heterogeneity is estimated on the combined treatment group and holdout control. The above estimates report the medians over 1000 such repetitions.

Appendix A: Sample Construction and Randomization Procedures

The study was conducted in 14 districts in five regions of Ghana: Northern, Upper East, Ashanti, Bono, and Bono East. In each district, IPA and Heifer International met with District Assembly staff (i.e. local government) to identify each community in the district, and to select communities that (a) had at least 50 compounds,[21] (b) were accessible by road from the district capital (to allow staff based in the district capital to travel to the communities), and (c) did not have programs similar to Heifer's graduation program already in operation.

In each community that fulfilled these initial criteria IPA administered a census of all households, in which we collected contact information and administered a proxy means test. In total 68,309 households in 366 communities were part of the census. Surveying took place in two waves, in January and February 2016 for Northern and Upper East regions, and from May through August, 2016 in Bono, Bono East and Ashanti.

Following the census, communities were deemed eligible if they had at least 45 compounds. We selected the 40 compounds with the lowest average household proxy means test score, and for each compound, randomly chose one household to include in our eligible sample, which consisted of 40 households in each community.[22]

For each of the intervention randomizations, we performed a re-randomization stratification procedure. We randomized a predetermined 10,000 times, tested for balance on a vector of characteristics (listed in Appendix Table 1) and picked the randomization with the maximum minimum p-value. This procedure was applied to both the community level randomizations and the within community randomizations.

For inclusion in the survey sample, we randomized which households in our eligible sample would be surveyed. This randomized sample selection took place after the community-level randomization because the number of households sampled per community depended on treatment status.

---

[21]A compound is a cluster of households living in separate dwellings clustered within a single structure.

[22]We only worked with one household per compound because of concerns about within-compound spillovers.

Appendix B: Description of Randomization Inference Procedure

The multi-stage nature of our randomization procedure (community-level randomization, followed by randomization to determine which households in pure control communities are included in the final sample, followed by randomization of the gender of CBT in a community, and individual level-randomization into CBT or control) motivates our use of randomization inference. We therefore follow the general procedure laid out by Young (2019), adapted to the specifics of our randomization procedure. In particular, we implemented the following procedure, for each of 2,000 simulations, (again following Young (2019), who finds "no appreciable change in rejection rates beyond 2,000 draws"):

1. Using community-level data obtained from the census, re-randomize 100 times to assign placebo treatments, test for balance on characteristics (listed in Appendix Table 1, Panel A), choose the randomization with the maximum minimum p-value.

2. Assign sample weights to households based on their placebo community assignment, reflecting the fact that a smaller number of households were included in the sample in pure control and CBT only communities in the true randomization assignment. For example, an household assigned to a pure control community in the real randomization (in which we randomly selected 17 of the 40 eligible households to include in our study) but was assigned to be in a full program community in the placebo randomization (in which all 40 eligible households were chosen) would be given a sample weight of (40/17).

3. Using household and adult-level data, re-randomize 100 times to assign placebo (i) CBT gender in a given community, and (ii) individual assignment into CBT or control, test for balance on characteristics (listed in Appendix Table 1, Panel C), choose the randomization with the maximum minimum p-value.

4. With the placebo treatment assignments, regress our outcome variables of interest on the placebo treatment, and in tests of heterogeneity, on the interaction between the true baseline outcome (distress, gender) and the placebo treatment. Store estimates of the coefficients of (i) average treatment effects, and for tests of heterogeneity, of (ii) sub-group treatment effects, and (iii) the difference in coefficient between the two sub-groups.

Our inference involves comparing the true point estimates (and in cases of heterogeneity, the difference in coefficients) to the empirical distribution of coefficients (differences) from our 2,000 simulations. Our "RI p-value" is equal to the share of the 2,000 simulations in which the absolute value of the coefficient (difference) is larger than the absolute

value (absolute value of the difference) of the results from our true randomization assignment.

We implement stages 1-3 (i.e. the placebo randomizations) separately for the "Northern Belt" (Northern and Upper East) and "Middle Belt" (Ashanti, Bono, Bono East) parts of our sample, reflecting the way in which we performed the actual randomization (we completed data collection activities and conducted randomizations for the Northern Belt before proceeding to the Middle Belt). For stage 4, the regression, we pool the full sample.

One aspect to note is that in our initial randomization procedure, we re-randomized a pre-specified 10,000 times to determine the maximum minimum p-value, where in these simulations we re-randomize a pre-specified 100 times. This adjustment was made for computational reasons, as given our nested randomizations we face the curse of dimensionality in exactly replicating our procedure.[23] We found when comparing simulations with 10,000 to 100 re-randomizations that our balance did not seem to differ appreciably; results available upon request.

---

[23]We estimated that exactly replicating the procedure would take approximately 160 days for the code to run.

Appendix C: Bandwidth Index Construction and Results

In our main text, we propose two hypotheses regarding the broad potential benefits to CBT in a low-income population: (1) the poor are especially vulnerable to subsequent mental distress, even if characterized as "well" at a given point in time, and (2) CBT helps individuals strengthen their socioemotional and cognitive skills, even in the absence of psychological distress.

To assess our first hypothesis, we test whether the benefits of CBT are concentrated among individuals whose baseline characteristics predict subsequent mental distress. We find that while we do not observe heterogeneity of treatment effects by baseline psychological distress, our we do observe heterogeneity by predicted endline distress, evidence consistent with our vulnerability hypothesis.

For completeness, we also test whether the "bandwidth" benefits to CBT are concentrated among individuals predicted to have low bandwidth at a subsequent point in time. To implement this procedure, we first build a measure of endline bandwidth, (a standardized index, with one half weight assigned to cognitive measures, and one half to socioemotional skills). We follow the same procedure as for predicted Kessler Score, outlined in Section 4.3, splitting our control sample in half, and using cross-validated LASSO to predict endline bandwidth for these individuals. Using the estimated regressors, we then estimate a "predicted bandwidth score" for both our holdout control sample and our treatment households, and test for an interaction between predicted bandwidth and receiving CBT. These results are reported in Appendix Table 8. In contrast to Table 4 of our main text, we are unable to reject no heterogeneity of treatment effects by predicted bandwidth. We do observe a positive relationship between predicted bandwidth and our indices, suggesting that health and economic outcomes are positively correlated with predicted socioemotional skills and cognition. However, our results do not provide evidence in favor of the concentrated impacts of CBT on bandwidth within a particular subset of the distribution (though note however, that our confidence intervals for these interactions are meaningfully larger than in Table 4).

Appendix D: Structured Ethics Appendix

For more explanation of each question, see Asiedu et al. 2020.

1. Policy Equipoise

Is there policy equipoise? That is, is there uncertainty regarding participants' net benefits from each arm of the study relative to the other arms and to the best possible policy to which participants could have access? If not, ethical randomization requires two conditions related to scarcity: (1) Was there scarcity, i.e., did the inclusion of multiple arms change the expected aggregate value of the programs delivered? (2) Do all ex-ante identifiable participants have equal moral or legal claims to the scarce programs?

If there is no reasonable expectation that one arm of the study produces more benefits to participants than any other arm or than the best possible alternative policy, then randomization is ethically unproblematic. If not, then excluding some participants from the superior treatment arm can only be justified by scarcity. Scarcity conditions are twofold: (1) resources are not sufficient, given constraints, to include all participants in the superior treatment arm; (2) no ex-ante identifiable participants are excluded from the superior arm and have a greater claim to those resources than any participant assigned to the superior arm. See MacKay 2018 for more complete discussions of policy equipoise.

The treatment arm provides group CBT therapy to a general population of the poor, rather than to individuals with a common identified mental health difficulty. There was no consensus among experts regarding the effectiveness of this form of CBT for a general population, so the control and treatment arms were in policy equipoise. Furthermore, for those in mental distress at the time of the intervention, we believe that there is equipoise given limited evidence of effectiveness in this setting and with CBT delivered in groups by lay counsellors. Regardless, should there not be equipoise, there was scarcity in that the program had a limited budget for delivering CBT to communities.

2. Role of researchers with respect to implementation

Are researchers "active" researchers, i.e. did the researchers have direct decision making power over whether and how to implement the program? If YES, what was the disclosure to participants and informed consent process for participation in the program? Providing IRB approval details may be sufficient but further clarification of any important issues should be discussed here. If NO, i.e., implementation was separate, explain the separation.

A researcher should be considered "active" if, for example, the implementing staff are employed by an institution at which the PI is employed, and the staff report either

directly or indirectly to the PI at this institution with regard to this project. Or if researchers control funding for implementation, or have direct decision-making power over key implementation decisions.

Some key factors that help illuminate whether the researchers are "active" or not (here "researchers" are defined as the PIs and the staff that report directly or indirectly to the PIs): Did researchers directly provide any of the interventions, or parts thereof, to participants? Did researchers interact directly with participants and implicitly endorse one or more of the interventions?

The research team played an active role in the design of the program, but the program was implemented by a third party. IRB approval was received from the University of Ghana Medical School, IPA, Yale University and Northwestern University. Informed consent from participants was limited to consent to take part in a survey, and not the intervention. Lack of informed consent for the intervention aspect is justified because of the voluntary nature of the intervention; the independent purpose of the intervention as a non-research service for those in the community; and, the fact that the participants were not a vulnerable population seeking advice from the research team.

3. Potential harms to participants or nonparticipants from the interventions or policies

Does the intervention, policy or product being studied pose potential harm to participants or non-participants? Related, are participants or likely affected non-participants particularly vulnerable? Also related, are participants' access to future services or policies changed because of participation in the study? If yes to any of the above, what is being done to mitigate such risks

It may be important to consider whether the researchers are "active" (see above) or not for this discussion. If the researchers are "active", then they are responsible for the potential harms, and thus a robust discussion is appropriate. If the researchers are not "active", then while they may not be responsible for potential harms, a discussion of this would be appropriate here.

There will almost always be some potential harms, if nothing else because of complementary investments such as time that participants in an intervention necessarily redirect from one activity to another. Quantifying these risks and complementary investments may be difficult ex-ante, but a discussion of what they are here would help the reader assess their likely importance relative to the potential benefits of the tested intervention. Also note that measuring any harms ex-post may be the exact reason for the study, particularly when the intervention is common.

If risks to nonparticipants exist, discuss the mechanisms through which the risk arises

from the study and provide an estimate of the magnitude of the risk and the probability of harm.

The IRB reviewed protocols for the CBT program, participation in which was voluntary and from which individuals were always free to withdraw. Protocols were in place for responding to sensitive issues and distress that emerged during or as a result of the sessions. In particular, anyone identified in surveys as in distress was directed to the community psychiatric nurse for help regardless of which arm they were randomized into.

The sessions did require participation, effort and time, but these costs were small in magnitude, and always under the control of the participants. Participants were not required to attend sessions, and there was no consequence to them for non-attendance

4. Potential harms to research participants or research staff from data collection (e.g., surveying, privacy, data management) or research protocols (e.g., random assignment)

Are data collection and/or research procedures adherent to privacy, confidentiality, risk-management, and informed consent protocols with regard to human subjects? Are they respectful of community norms, e.g., community consent not merely individual consent, when appropriate? Are there potential harms to research staff from conducting the data collection that are beyond "normal" risks?

Example of sub-questions to consider as part of the broad question: Are there any risks that could ensue because of the data collection process or storage, e.g. discomfort to being asked certain questions or breach of confidentiality? If so, what are the mitigation strategies? Are there costs to the participant for the data collection process, such as their time, and if so, what is the strategy or rationale for offsetting this cost?

Because these are all issues covered by most IRB processes, a sufficient explanation for a "yes" response may be to provide the IRB approval numbers for all IRBs that have approved the project. However, if there are particular issues that are important to discuss, please do so here.

Harms to research staff could include, e.g., exposure to political violence, exposure to unusual levels of a communicable disease, mistrust due to lack of perceived lack of community consent, or emotional wellbeing from surveying about difficult subject matters. This would not include, e.g., traffic accidents.

Data collection procedures were in adherence with human subjects protocols and respectful of community norms. There were no special risks to research staff.

5. Financial and reputational conflicts of interest Do any of the researchers have financial conflicts of interest with regard to the results of the research? Do any of the researchers have potential reputational conflicts of interest?

We define financial conflicts of interest as that used by the researcher's institutional (e.g., their university) guidelines. We define a reputational conflict of interest as one in which prior writing or advocacy could be contradicted by specific results pursued in this study, and such contradiction would pose reputational risks to the author.

None.

6. Intellectual freedom

Were there any contractual limitations on the ability of the researchers to report the results of the study? If so, what were those restrictions, and who were they from?

This could include, for example, approval of release of the paper and restrictions on data release, but does not include things such as a "comment period" during which interested parties have a right to review and provide comments prior to release but not to control the outputs of the study.

No restrictions.

7. Feedback to participants or communities

Is there a plan for providing feedback on research results to participants or communities? If yes, what is the plan? If not, why not?

Engaging in post-study feedback is a way of acknowledging the agency of participants and communities, and is thus a desired practice. However, it may be impractical due to costs, timing, challenges communicating the results, or potential harms if such communication may itself change behavior in undesirable ways.

We hope to provide feedback as part of the closing procedure for the overall Escaping Poverty research program, of which this is part.

8. Foreseeable misuse of research results

Is there a foreseeable and plausible risk that the results of the research will be misused and/or deliberately misinterpreted by interested parties to the detriment of other interested parties? If yes, please explain any efforts to mitigate such risk.

In settings with strong imbalances of power between interested parties, there may be foreseeable risks that a powerful party could use deliberately selected research findings to

their advantage and to the harm of participants or non-participants, including for general public policy. For example, if the research might reveal the vulnerability of some that can be exploited for the gain of the more powerful party, what steps does the researcher plan to mitigate this risk?

None.

9. Other Ethics Issues to Discuss

None.

Appendix Tables

Appendix Table 1: Variables Used in Re-Randomization Procedures

Panel A: Variables in Re-Randomization to Determine Community-Level Assignment

District-level dummies
Mean proxy means test score
SD of proxy means tests in community
Paved road connected to village
Electricity in village
Distance from nearest market
Number of compounds in community

Panel B: Variables in Re-Randomization to Determine Final Sample of Households

Male head of household
Number of co-resident co-wives
Proxy means test score
Age of household head
Average proxy means score among HHs in compound
Number of households in compound

Panel C: Variables used in Re-Randomization to Determine CBT Treatment Assignment

Age of household head
Number of children under 5
Household size
Cash savings balance, winsorized
Land owned, winsorized
Business profits, winsorized
Any adult skipped meals last month
Total asset value, winsorized
Total livestock value, winsorized
Kessler Score, baseline
Missing Kessler Score, baseline
No male head of household present

This table lists the variables used in our re-randomization procedure to determine (A) whether a community is pure control, pure CBT, or full program, (B) which households in pure control and pure CBT communities to sample and include in our study, and (C) which individuals in pure CBT or full program communities were offered the CBT program

Appendix Table 2: Average Treatment Effects on Mental Health, by Control Group Definition

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Mental Health Index	Kessler Score	No distress (Kessler < 20)	No moderate or severe distress (Kessler < 25)	No severe distress (Kessler <30)	Mental Health Self Rating (1/4)	Days in month without poor mental health
Panel A: Treatment vs All Control							
Assigned to CBT	0.15	-1.36	0.06	0.06	0.04	0.07	0.53
	(0.029)	(0.268)	(0.017)	(0.016)	(0.012)	(0.029)	(0.249)
Observations	7,445	7,439	7,439	7,439	7,439	7,445	7,412
R-squared	0.09	0.09	0.06	0.05	0.03	0.04	0.03
Control mean	0.00	21.4	0.4	0.7	0.8	2.8	25.3
Panel B: Treatment vs Pure Control							
Assigned to CBT	0.11	-1.17	0.052	0.054	0.031	0.059	0.27
	(0.044)	(0.382)	(0.023)	(0.023)	(0.016)	(0.045)	(0.336)
Observations	3,601	3,601	3,601	3,601	3,601	3,601	3,589
R-squared	0.09	0.09	0.06	0.05	0.03	0.04	0.03
Control mean	0.01	21.6	0.4	0.7	0.8	2.9	25.2
Panel C: Spillover Treatment vs Spillover Control							
Assigned to CBT	-0.06	0.31	-0.007	-0.017	-0.012	-0.013	-0.57
	(0.046)	(0.406)	(0.022)	(0.023)	(0.017)	(0.043)	(0.347)
Observations	6,155	6,149	6,149	6,149	6,149	6,155	6,127
R-squared	0.10	0.09	0.06	0.05	0.03	0.04	0.03
Control mean	0.01	21.6	0.4	0.7	0.8	2.9	25.2

Panel A presents the same results in our main analysis, in which we include all control individuals (both individuals in control villages, and control individuals in pure CBT and full program communities). Panel B restricts the control group to individuals in control villages (i.e. fully eliminating the possibility of within-village spillovers, at the cost of a reduced sample). Panel C tests for in-village spillovers, by comparing individuals in pure CBT or full program communities who did not receive the program to individuals in pure control communities. In all specifications, standard errors are clustered at the village level.

Appendix Table 3: Average Treatment Effects on Physical Health, by Control Group Definition

	(1)	(2)	(3)
	Physical Health Index	Physical Health Self Rating (1/4)	Days in month without poor physical health
Panel A: Treatment vs All Control			
Assigned to CBT	0.15	0.12	0.89
	(0.027)	(0.024)	(0.227)
Observations	7,445	7,445	7,416
R-squared	0.12	0.10	0.06
Control mean	0.00	3.03	25.57
Panel B: Treatment vs Pure Control			
Assigned to CBT	0.12	0.09	0.74
	(0.034)	(0.031)	(0.254)
Observations	3,601	3,601	3,593
R-squared	0.10	0.10	0.05
Control mean	0.00	3.05	25.45
Panel C: Spillover Treatment vs Spillover Control			
Assigned to CBT	-0.05	-0.04	-0.28
	(0.031)	(0.030)	(0.236)
Observations	6,155	6,155	6,130
R-squared	0.13	0.10	0.07
Control mean	0.00	3.05	25.45

Panel A presents the same results in our main analysis, in which we include all control individuals (both individuals in control villages, and control individuals in pure CBT and full program communities). Panel B restricts the control group to individuals in control villages (i.e. fully eliminating the possibility of within-village spillovers, at the cost of a reduced sample). Panel C tests for in-village spillovers, by comparing individuals in pure CBT or full program communities who did not receive the program to individuals in pure control communities. In all specifications, standard errors are clustered at the village level.

Appendix Table 4: Average Treatment Effects on Socio-Emotional Skills, by Control Group Definition

	(1)	(2)	(3)	(4)
	Socio-Emotional Skill Index	Generalized Self-Efficacy Score	Grit Score	Self-Control Score
Panel A: Treatment vs All Control				
Assigned to CBT	0.27	0.29	0.19	0.12
	(0.035)	(0.033)	(0.033)	(0.036)
Observations	7,444	7,444	7,441	7,436
R-squared	0.09	0.06	0.05	0.07
Control mean	0.00	0.00	0.00	0.00
Panel B: Treatment vs Pure Control				
Assigned to CBT	0.33	0.39	0.20	0.15
	(0.052)	(0.045)	(0.050)	(0.051)
Observations	3,601	3,601	3,601	3,601
R-squared	0.10	0.08	0.06	0.07
Control mean	0.06	0.07	0.02	0.03
Panel C: Spillover Treatment vs Spillover Control				
Assigned to CBT	0.10	0.16	0.01	0.06
	(0.050)	(0.044)	(0.048)	(0.046)
Observations	6,154	6,154	6,151	6,146
R-squared	0.08	0.06	0.05	0.07
Control mean	0.06	0.07	0.02	0.03

Panel A presents the same results in our main analysis, in which we include all control individuals (both individuals in control villages, and control individuals in pure CBT and full program communities). Panel B restricts the control group to individuals in control villages (i.e. fully eliminating the possibility of within-village spillovers, at the cost of a reduced sample). Panel C tests for in-village spillovers, by comparing individuals in pure CBT or full program communities who did not receive the program to individuals in pure control communities. In all specifications, standard errors are clustered at the village level.

	(1)	(2)	(3)	(4)	(5)
	Cognition Index	Raven's Progressive Matrices, Indexed	Digit Span: Forwards, Indexed	Digit Span: Backwards, Indexed	Executive Function Test, Indexed
Panel A: Treatment vs All Control					
Assigned to CBT	0.08	0.03	0.076	0.070	0.053
	(0.034)	(0.034)	(0.032)	(0.031)	(0.034)
Observations	7,445	7,439	7,440	7,440	7,445
R-squared	0.09	0.06	0.05	0.07	0.00
Control mean	0.00	0.00	0.00	0.00	0.00
Panel B: Treatment vs Pure Control					
Assigned to CBT	0.085	0.007	0.082	0.095	0.055
	(0.046)	(0.054)	(0.045)	(0.039)	(0.044)
Observations	3,601	3,599	3,601	3,601	3,601
R-squared	0.10	0.08	0.06	0.07	0.00
Control mean	0.06	0.07	0.02	0.03	0.00
Panel C: Spillover Treatment vs Spillover Control					
Assigned to CBT	0.00	-0.04	0.015	0.033	-0.001
	(0.039)	(0.048)	(0.042)	(0.032)	(0.036)
Observations	6,155	6,149	6,150	6,150	6,155
R-squared	0.08	0.06	0.05	0.07	0.00
Control mean	0.06	0.07	0.02	0.03	0.00

Panel A presents the same results in our main analysis, in which we include all control individuals (both individuals in control villages, and control individuals in pure CBT and full program communities). Panel B restricts the control group to individuals in control villages (i.e. fully eliminating the possibility of within-village spillovers, at the cost of a reduced sample). Panel C tests for in-village spillovers, by comparing individuals in pure CBT or full program communities who did not receive the program to individuals in pure control communities. In all specifications, standard errors are clustered at the village level.

Appendix Table 6: Average Treatment Effects on Economic Outcomes, by Control Group Definition

	(1)	(2)	(3)	(4)
	Economic Index	Days in which poor mental or physical health did not keep individual from doing regular activities	Self-Reported Economic Status (1/10)	Projected Economic Status in 5 years (1/10)
Panel A: Treatment vs All Control				
Assigned to CBT	0.19	0.37	0.43	0.35
	(0.035)	(0.222)	(0.075)	(0.100)
Observations	7,445	7,396	7,445	7,445
R-squared	0.07	0.04	0.05	0.05
Control mean	0.00	26.86	3.08	5.79
Panel B: Treatment vs Pure Control				
Assigned to CBT	0.19	0.30	0.50	0.33
	(0.048)	(0.267)	(0.105)	(0.142)
Observations	3,601	3,583	3,601	3,601
R-squared	0.08	0.04	0.06	0.06
Control mean	-0.03	26.67	3.04	5.80
Panel C: Spillover Treatment vs Spillover Control				
Assigned to CBT	0.00	-0.13	0.11	-0.04
	(0.046)	(0.244)	(0.092)	(0.136)
Observations	6,155	6,114	6,155	6,155
R-squared	0.06	0.05	0.04	0.05
Control mean	-0.03	26.67	3.04	5.80

Panel A presents the same results in our main analysis, in which we include all control individuals (both individuals in control villages, and control individuals in pure CBT and full program communities). Panel B restricts the control group to individuals in control villages (i.e. fully eliminating the possibility of within-village spillovers, at the cost of a reduced sample). Panel C tests for in-village spillovers, by comparing individuals in pure CBT or full program communities who did not receive the program to individuals in pure control communities. In all specifications, standard errors are clustered at the village level.

Appendix Table 7a: Variables selected by LASSO Procedure on Predicted Kessler Score, Part I

Variable	Share of repetitions in which chosen by LASSO	Mean Beta in LASSO Predicting Kessler Score
During the last 30 days, about how often did you feel depressed?	0.950	0.370
Mental Health Index	0.943	-0.411
Physical Health Index	0.918	-0.380
Female	0.902	0.580
Have experienced the threat of death or serious bodily harm	0.850	0.511
Asset Value, Winsorized fraction .01, high only	0.834	-0.0002
Kessler Depression Score	0.832	0.081
witnessed family violence	0.779	0.441
During the last 30 days, about how often did you feel nervous?	0.767	0.256
Have experienced a life-threatening illness	0.756	0.407
Cantril Ladder score (1/10)	0.716	-0.194
I have been obsessed with a certain idea or project for a short time but later lost interest	0.701	-0.136
I finish whatever I begin.	0.696	-0.225
I often set a goal but later choose to pursue a different one.	0.690	-0.129
Business revenue, Winsorized fraction .01, high only	0.660	-0.001
Have experienced a sudden death of a loved one	0.639	0.442
Forwards Digits Remembered	0.634	0.150
Have experienced childhood physical abuse	0.577	0.355
Total value of savings, Winsorized fraction .01, high only	0.556	-0.0004
I say inappropriate things.	0.542	-0.141
During the last 30 days, about how often did you feel worthless	0.502	0.158
During the past 30 days, for about how many days did poor physical or mental health…	0.459	0.029
In general, how would you rate your physical health? higher is better	0.458	-0.250
Do you consider yourself a member of a political party?	0.432	-0.302
Setbacks don't discourage me.	0.402	0.090
During the last 30 days, about how often did you feel restless or fidgety?	0.396	0.192
During the last 30 days, about how often did you feel so sad that nothing could calm you?	0.381	0.163
Land size, Winsorized fraction .01, high only	0.378	-0.015
Have experienced a life-threatening or permanently disabling event for loved one	0.364	0.284
I have overcome setbacks to conquer an important challenge.	0.357	0.117
Backwards Digits Remembered	0.345	-0.154
Days in month without poor mental health	0.327	-0.024
I have a hard time breaking bad habits.	0.327	0.113
Have you met with your assemblyman at any time in the last 12 months?	0.303	-0.262
I am confident that I could deal efficiently with unexpected events.	0.269	-0.132
Sometimes I can't stop myself from doing something, even if I know it's wrong.	0.267	-0.110
In general, how would you rate your mental health? higher is better	0.255	-0.213
During the last 30 days, about how often did you feel helpless?	0.245	0.131
During the last 30 days, about how often did you feel so restless that you could not sit still?	0.239	0.153
I am diligent.	0.226	-0.140
People would say that I have very strong self-discipline	0.199	0.126
Household Head age	0.196	0.009
I have difficulty maintaining my focus on projects that take more than a few months	0.190	0.080

Appendix Table 7b: Variables selected by LASSO Procedure on Predicted Kessler Score, Part II

Variable	Share of repetitions in which chosen by LASSO	Mean Beta in LASSO Predicting Kessler Score
I'm good at resisting temptation.	0.179	0.115
Have experienced warfare or combat	0.176	0.230
I do things that feel good in the moment but regret later on.	0.170	-0.064
I refuse things that are bad for me, even if they are fun.	0.165	0.068
During the last 30 days, about how often did you feel that everything was an effort	0.163	0.085
Where would you be on this ladder compared to the rest of the people of Ghana?	0.163	-0.102
Have experienced a robbery involving a weapon	0.163	-0.392
I have achieved a task that took years of work.	0.156	-0.083
Have experienced another type of accident	0.153	0.252
EC Number Correct	0.145	-0.031
New ideas and projects sometimes distract me from previous ones.	0.142	-0.064
Have witnessed a severe assault of an acquaintance or stranger	0.139	0.182
I can always manage to solve difficult problems if I try hard enough.	0.135	0.166
During the last 30 days, about how often did you feel tired for no reason?	0.133	0.069
Days in month without poor physical health	0.124	-0.021
I often act without thinking through all the alternatives.	0.109	-0.036
I am a hard worker.	0.107	-0.094
I can solve most problems if I invest the necessary effort.	0.104	0.102
On the same ladder, where do you expect to be in five years?, z-score	0.097	0.123
I can remain calm when facing difficulties because I can rely on my coping abilities	0.092	-0.041
experienced a severe assault by an acquaintance or stranger	0.092	-0.049
I get distracted easily.	0.089	0.050
If I am in trouble, I can usually think of a solution.	0.086	-0.008
experienced a motor vehicle accident	0.085	0.131
If someone opposes me, I can find the means and ways to get what I want.	0.084	-0.047
Pleasure and fun sometimes keep me from getting work done.	0.084	0.037
Raven's Index Score	0.083	-0.013
number of household members under age 5	0.083	-0.019
number of household members	0.080	-0.017
Have you met with your chief at any time in the last 12 months?	0.078	-0.078
I can usually handle whatever comes my way.	0.078	0.035
When I am confronted with a problem, I can usually find several solutions.	0.076	-0.062
It is easy for me to stick to my aims and accomplish my goals.	0.074	0.048
On the same ladder, where do you expect to be in five years?, z-score	0.065	0.037
Thanks to my resourcefulness, I know how to handle unforeseen situations.	0.050	0.047
Grit index z-score	0.045	-0.089
During the last 30 days, about how often did you feel so nervous that nothing could calm you?	0.036	-0.143
GSES self-efficacy z-score	0.012	-0.051
Self control z-score	0.006	0.092
Livestock Value, Winsorized fraction .01, high only	0.001	-0.0001

Appendix Table 8: Heterogeneity by Predicted Bandwidth Score

	(1)	(2)	(3)	(4)	(5)	(6)
	Kessler Psychological Distress Score	Mental Health Index	Physical Health Index	Economic outcomes Index	Socioemotional Skills Index	Cognition Index
Assigned to CBT	-2.245	0.235	0.246	0.272	0.257	0.028
	[-3.491, -1.005]	[0.075, 0.395]	[0.109, 0.383]	[0.078, 0.466]	[0.082, 0.432]	[-0.134, 0.190]
Unadjusted p-value	0.003	0.016	0.004	0.021	0.016	0.769
P-Value adjusted for sample split	0.006	0.032	0.008	0.042	0.032	1.000
Predicted Bandwidth Index from Baseline Covariates						
(rescaled to minimum area of common support)	-2.236	0.255	0.186	0.155	0.539	1.201
	[-2.699, -1.776]	[0.195, 0.315]	[0.126, 0.246]	[0.094, 0.216]	[0.478, 0.599]	[1.141, 1.261]
Unadjusted p-value	0.000	0.000	0.000	0.000	0.000	0.000
P-Value adjusted for sample split	0.000	0.000	0.000	0.000	0.000	0.000
Assigned to CBT x Predicted Bandwidth Index	-0.546	0.051	0.064	0.053	0.013	-0.046
	[-1.439, 0.348]	[-0.063, 0.164]	[-0.041, 0.167]	[-0.087, 0.191]	[-0.111, 0.137]	[-0.160, 0.068]
Unadjusted p-value	0.316	0.464	0.316	0.537	0.815	0.503
P-Value adjusted for sample split	0.632	0.928	0.632	1.000	1.000	1.000
Estimated treatment effect: individual with 10th percentile predicted Kessler Score	-1.225	0.141	0.127	0.176	0.232	0.114
Estimated treatment effect: individual with 90th percentile predicted Kessler Score	-1.817	0.195	0.196	0.231	0.246	0.064

In each simulation, the control group is divided in two groups: a training control sample and holdout control. The LASSO is estimated using k-fold estimation on the training control, predicting endline Bandwidth index (Panel B) using baseline covariates. The above specification to estimate treatment effect heterogeneity is estimated on the combined treatment group and holdout control. The above estimates report the medians over 1000 such simulations.

Appendix Table 9a: Variables selected by LASSO Procedure on Predicted Bandwidth Index, Part I

Variable	Share of repetitions in which chosen by LASSO	Mean Beta in LASSO Predicting Bandwidth Index
Female	1.000	-0.165
Backwards Digits Remembered	1.000	0.162
Forwards Digits Remembered	1.000	0.058
Raven's Index Score	1.000	0.053
EC Number Correct	0.990	0.017
Have you met with your assemblyman at any time in the last 12 months?	0.989	0.099
experienced a motor vehicle accident	0.915	0.072
Self control z-score	0.912	0.060
I say inappropriate things.	0.852	0.030
Business revenue, Winsorized fraction .01, high only	0.852	0.0002
During the last 30 days, about how often did you feel that everything was an effort	0.834	-0.024
Household Head age	0.807	-0.002
Have experienced another type of accident	0.765	0.057
On the same ladder, where do you expect to be in five years?, z-score	0.750	0.023
I am a hard worker.	0.708	0.026
I do things that feel good in the moment but regret later on.	0.696	0.019
I have been obsessed with a certain idea or project for a short time but later lost interest	0.691	0.016
I have a hard time breaking bad habits.	0.659	0.019
Grit index z-score	0.585	0.023
Pleasure and fun sometimes keep me from getting work done.	0.573	0.016
I can usually handle whatever comes my way.	0.565	0.021
Have experienced childhood physical abuse	0.558	-0.042
People would say that I have very strong self-discipline	0.552	0.018
On the same ladder, where do you expect to be in five years?, z-score	0.545	0.019
Have experienced a sudden death of a loved one	0.521	0.044
I am confident that I could deal efficiently with unexpected events.	0.490	0.018
experienced a severe assault by an acquaintance or stranger	0.470	-0.048
During the last 30 days, about how often did you feel restless or fidgety?	0.462	-0.017
During the last 30 days, about how often did you feel depressed?	0.411	-0.015
Do you consider yourself a member of a political party?	0.406	0.031
I can remain calm when facing difficulties because I can rely on my coping abilities	0.363	0.017
In general, how would you rate your physical health? higher is better	0.358	0.020
Have experienced a life-threatening illness	0.354	0.036
I can always manage to solve difficult problems if I try hard enough.	0.348	0.019
Days in month without poor mental health	0.305	-0.002
Have experienced a life-threatening or permanently disabling event for loved one	0.304	-0.033
I have overcome setbacks to conquer an important challenge.	0.270	0.011
During the last 30 days, about how often did you feel so restless that you could not sit still?	0.266	-0.015
Thanks to my resourcefulness, I know how to handle unforeseen situations.	0.263	0.016
New ideas and projects sometimes distract me from previous ones.	0.258	0.011
Land size, Winsorized fraction .01, high only	0.251	0.002
Have experienced the threat of death or serious bodily harm	0.247	-0.034
Have experienced a robbery involving a weapon	0.242	0.048

Appendix Table 9b: Variables selected by LASSO Procedure on Predicted Bandwidth Index, Part II

Variable	Share of repetitions in which chosen by LASSO	Mean Beta in LASSO Predicting Bandwidth Index
When I am confronted with a problem, I can usually find several solutions.	0.241	0.017
During the last 30 days, about how often did you feel so sad that nothing could calm you?	0.239	-0.012
number of household members	0.214	-0.004
I get distracted easily.	0.211	0.010
Have you met with your chief at any time in the last 12 months?	0.203	-0.031
Have witnessed a severe assault of an acquaintance or stranger	0.200	0.023
Setbacks don't discourage me.	0.177	-0.010
GSES self-efficacy z-score	0.166	0.016
Have experienced warfare or combat	0.159	-0.024
I finish whatever I begin.	0.157	0.011
I often set a goal but later choose to pursue a different one.	0.156	0.008
Cantril Ladder score (1/10)	0.155	0.011
In general, how would you rate your mental health? higher is better	0.149	0.012
I'm good at resisting temptation.	0.146	-0.016
witnessed family violence	0.146	-0.017
If I am in trouble, I can usually think of a solution.	0.129	0.011
During the past 30 days, for about how many days did poor physical or mental health…	0.125	-0.002
During the last 30 days, about how often did you feel worthless?	0.123	-0.009
I can solve most problems if I invest the necessary effort.	0.121	-0.022
If someone opposes me, I can find the means and ways to get what I want.	0.120	0.007
I have difficulty maintaining my focus on projects that take more than a few months	0.118	0.007
During the last 30 days, about how often did you feel so nervous that nothing could calm you down	0.104	0.013
I am diligent.	0.095	0.007
I have achieved a task that took years of work.	0.090	0.005
During the last 30 days, about how often did you feel tired for no reason?	0.089	0.007
It is easy for me to stick to my aims and accomplish my goals.	0.087	-0.007
During the last 30 days, about how often did you feel nervous?	0.087	-0.006
Where would you be on this ladder compared to the rest of the people of Ghana?	0.078	-0.005
Sometimes I can't stop myself from doing something, even if I know it's wrong	0.074	0.005
I refuse things that are bad for me, even if they are fun.	0.072	0.007
During the last 30 days, about how often did you feel helpless?	0.069	0.007
Physical Health Index	0.069	0.011
Days in month without poor physical health	0.065	-0.001
number of household members under age 5	0.060	0.003
Kessler Depression Score	0.052	-0.002
I often act without thinking through all the alternatives.	0.051	-0.001
Mental Health Index	0.037	-0.011
Asset Value, Winsorized fraction .01, high only	0.000	-
Livestock Value, Winsorized fraction .01, high only	0.000	-
Total value of savings, Winsorized fraction .01, high only	0.000	-