

Journal of Development Economics

Using Group Interpersonal Psychotherapy to Improve the Well-Being of Adolescent Girls

--Manuscript Draft--

Manuscript Number:	DEVEC_2019_1286R2
Article Type:	Registered Report Stage 1: Proposal
Section/Category:	Health, Education, gender, poverty
Keywords:	Adolescence; group interpersonal psychotherapy; Gender; Mental Health; Cash Transfers
Corresponding Author:	Sarah Baird George Washington University Milken Institute of Public Health Washington, DC UNITED STATES
First Author:	Sarah Baird
Order of Authors:	Sarah Baird Berk Ozler Chiara Dell'Aira Danish Us Salam
Abstract:	Adolescence is a time of heightened psychosocial vulnerability: the age of onset for most mental disorders likely to persist into adult life falls within ages 12–24 and neuropsychiatric disorders are now the leading cause of disability in adolescence. However, evidence is scant on what interventions can break the vicious cycle of poverty and poor mental health. Cognitive behavioral therapy and interpersonal psychotherapy (IPT) have been identified as well-established treatments for depression. We designed a three-arm cluster randomized controlled trial, which aims to (a) provide proof of concept for a low-cost and scalable group-based IPT to sustainably reduce depression among adolescent females in Uganda; (b) test the hypothesis that small cash transfers attached to psychotherapy can lead to larger reductions in depression in the longer-run than therapy alone; and (c) test the hypothesis that reducing depression during adolescence can translate into improved human capital accumulation.
Response to Reviewers:	

Sarah Baird
Associate Professor
George Washington University
Email: sbaird@email.gwu.edu

Re: Re-Submission to *JDE Registered Reports Stage 1*: “Using Group Interpersonal Psychotherapy to Improve the Well-Being of Adolescent Girls”

Dear Editor,

Please find attached a revised submission for the JDE Registered Report Stage 1, titled “*Using Group Interpersonal Psychotherapy to Improve the Well-Being of Adolescent Girls?*” by Sarah Baird, Berk Özler, Chiara Dell’Aira, and Danish Us Salam.

We have edited the submission for your four minor comments.

We look forward to hearing your reactions to this updated report.

Sincerely,

Sarah Baird

Ref: DEVEC 2019 1286R1

Title: Using Group Interpersonal Psychotherapy to Improve the Well-Being of Adolescent Girls

We thank the editor for his comments on our manuscript. We have addressed the final four minor tweaks. Below are the editors' comments (in **bold**) and our brief description as to how we responded to each of them.

Dear Dr. Baird,

Thank you for submitting your manuscript to Journal of Development Economics.

Awesome edits, thanks.

Great, thanks.

Four fairly minor tweaks please.

1) weird formatting in the pdf, see the intro area where there are some symbols instead of i think dashes.

Hmm. I think this may be a Mac issue. They look fine in the word version. I have managed to fix the two I saw in the introduction, and there are none in the word document. Apologies if any still remained in the PDF.

2) p6: change "statistically insignificant" to "not statistically significant"

Changed.

3) Need to opine more about the lack of the cash-only arm. explain why it'd be ideal, what it would be able to do that you can't now do, and why it isn't there.

We have now included the answer to why a 'cash-only' arm is not there in Section 2.2 (Interventions) and why it would be ideal and what it prevents us from doing, but also why the design is still adequate in Section 4 (Interpreting Results).

4) the mechanisms (beliefs, aspirations, eg) need to be in the set of 3 questions that start on page 5 (and elsewhere). so either combine this in as psychosocial first-stage changes and make this part of the first question, or separate it out into its own question. seems natural to combine with what you are putting into question #1 (depression) and just expand that, but up to you.

This was previously incorporated in research question 3. We have kept it here but expanded to make it more explicit. These mechanisms are really the short term pre-cursors to future human capital gains, so we think it fits best here.

Journal of Development Economics Registered Report Stage 1: Proposal
**Using Group Interpersonal Psychotherapy to
Improve the Well-Being of Adolescent
Girls**

3/17/2020

JEL codes: O10, C93, I15, J16, J13

Keywords: adolescence, mental health, group interpersonal psychotherapy, cash transfers, gender

Study pre-registration: This study is registered in the AEA RCT Registry and the unique identifying number is AEARCTR-0004208: <http://www.socialscienceregistry.org/trials/4208>.

It is also registered at ClinicalTrials.gov and has the unique identifying number NCT03966833: <https://clinicaltrials.gov/ct2/show/NCT03966833?term=NCT03966833&rank=1>.

The trial was registered in both registries prior to baseline data collection.

Abstract

Adolescence is a time of heightened *psychosocial vulnerability*: the age of onset for most mental disorders likely to persist into adult life falls within ages 12–24 and neuropsychiatric disorders are now the leading cause of disability in adolescence. However, evidence is scant on what interventions can break the vicious cycle of poverty and poor mental health. Cognitive behavioral therapy and interpersonal psychotherapy (IPT) have been identified as well-established treatments for depression. We designed a three-arm cluster randomized controlled trial, which aims to (a) provide proof of concept for a low-cost and scalable group-based IPT to sustainably reduce depression among adolescent females in Uganda; (b) test the hypothesis that small cash transfers attached to psychotherapy can lead to larger reductions in depression in the longer-run than therapy alone; and (c) test the hypothesis that reducing depression during adolescence can translate into improved human capital accumulation.

Proposed Timeline

Timeline

Activities	Period
Selection of BRAC branches and ELA clubs	April 2019
Screening and baseline	May 2019 - August 2019
Randomization (T/C)	August 2019
Program implementation	September 2019 – December 2019
Rapid follow up	December 2019 – January 2020
12-month follow up	May 2020 - August 2020
24-month follow up	May 2021 - August 2021

The above table provides a summary of the study’s main timeline. Prior to the initiation of data collection activities, our implementing partner BRAC Uganda identified 106 “active” ELA clubs within the Central Region of Uganda that were verified to offer regular daily sessions consistently delivered by the same BRAC-trained mentor. We selected all of these clubs for the study. Screening and baseline activities took place in communities surrounding the selected clubs (within a 0.5km radius). We first implemented a listing survey in order to identify female adolescents within the age bracket (13-19) living in the area. All young women for whom relevant consent and/or assent was obtained were then screened and, if found eligible, recruited into the study and interviewed as part of the baseline data collection.¹

Baseline data collection was followed by randomization of each cluster into treatment and control and then shortly thereafter program implementation began. At the end of the 14-week-long intervention, all study participants are being administered a rapid assessment survey to determine immediate impacts of the program on the primary outcomes.² Finally, two additional follow up surveys will be carried out 12 and 24 months after baseline.

¹ Please refer to “Section 2.5.1: Sampling strategy” for details on the listing area surrounding each ELA club, exact eligibility criteria, and the two-stage listing and sampling strategy.

² The principal investigators are blinded to these data while the registered report is under review.

1. Introduction

Mental and behavioral disorders are a main contributor to years lived with disability (YLDs), with major depressive and anxiety disorders being among the leading specific causes (Vos et al. 2012). Adolescence is also a time of heightened *psychosocial vulnerability*: the age of onset for most mental disorders likely to persist into adult life falls within ages 12–24 and neuropsychiatric disorders are now the leading cause of disability in adolescence (Patel et al. 2007; World Health Organization, or WHO, 2014). Poor mental health exacts a heavy toll on young women in Sub-Saharan Africa: along with HIV/AIDS and abortion, depression is one of the leading contributors to YLDs.

Mental health disorders developed during adolescence are of immediate intrinsic importance, and the World Bank and the WHO have a stated commitment to make mental health a global development priority (Vos et al. 2016). However, these disorders can also have negative long-run consequences through increased risky decision making (Di Clemente et al. 2001; Fishbein et al. 2006); future mental and physical health problems (Evans et al. 2007; McLoyd et al. 2009); lower educational achievement (Currie and Stabile 2006; Eisenberg and Golberstein 2009; Fletcher 2008, 2010; Fletcher and Wolfe 2008; Kessler et al. 1995; Stein and Kean 2000); and low self-esteem and self-efficacy.³

As poverty can also precipitate poor mental health (Dercon and Krishnan 2009; Lund et al. 2011; Patel et al. 2007), there may exist a vicious circle of poor mental health and low socioeconomic status (Patel and Kleinman 2003). Such a relationship makes it pertinent to identify interventions that can break this cycle – especially in low- and middle-income countries (LMICs). Lund et al. (2011) provides a review of interventions to improve mental health, such as residential drug treatment programs or psychotherapy, and finds that they were associated with improved economic outcomes. More recently, Baranov et al. (2018) reports results from a cluster-randomized trial of providing cognitive behavioral therapy (CBT) to perinatally depressed women in rural Pakistan and show that the psychotherapy not only substantially reduced the likelihood of suffering from postpartum depression after one year, but also had sustained effects on severity of depression, financial empowerment, and parental investments in children seven years after baseline. Evidence on effective interventions remains relatively scarce, however, especially those targeted to adolescents. For example, Patel et al. (2007) state: “We were unable to identify a single

³ Low self-esteem and self-efficacy, which are highly correlated with psychological distress among adolescents (Tait, French, and Hulse 2003), have also been shown to strongly influence schooling decisions, wages, and a host of behaviors that determine social and economic success (Heckman, Stixrud, and Urzua 2006; Krishnan and Krutikova 2012).

intervention targeted at young people in low-income and middle-income countries that improved mental-health outcomes.”

Important advances have been made in treatment development for adolescents, with recent reviews of psychosocial treatments for depression having identified CBT as the most thoroughly tested intervention that appears to meet the criteria for a well-established treatment, with interpersonal psychotherapy (IPT) also receiving considerable attention (Rosselló, Bernal, and Rivera-Medina 2008). CBT holds the idea that depression results from maladaptive information processing strategies and is maintained by dysfunctional behavioral responses. The therapy focuses on identifying and changing the function, content, and structure of cognitions associated with negative emotions or poor self-concept, then trying to teach the patient alternative methods of thinking or behaving. While IPT is time-limited, symptom-targeted, and is structured just like CBT, it focuses on interpersonal disputes, role transitions, grief, and interpersonal deficits – with the underlying hypothesis being that if the patient can solve the interpersonal problem or is able to change the relation to this problem, the depressive symptoms should resolve as well (Zhou et al. 2017).

Antidepressants have also been shown to be effective in reducing symptoms of acute depression in developing countries but cost and supply problems limit their feasibility. In Uganda, our study setting, IPT, with its focus on interpersonal relationships, has been shown to be compatible with Ugandan culture and efficacious in reducing depression and dysfunction (Bolton et al. 2003). Building on this foundation, StrongMinds – the international non-governmental organization that is the implementer of the group-based IPT (IPT-G) being evaluated here, has treated more than 25,000 women in Uganda since 2014 (<https://strongminds.org/strongminds-faqs/>). A primary motivation of this study is to ascertain whether IPT-G can also be effective in reducing depression and anxiety among adolescent females in Uganda.

As considerations of cost-effectiveness are more binding in low-resource settings, group-based IPT is attractive as an alternative treatment option to individual IPT. An important question is whether group-based therapy (for CBT or IPT) can be as effective as individual therapy. Rosselló, Bernal, and Rivera-Medina (2008) tested the group and individual formats of CBT and IPT among adolescents drawn from schools in San Juan, Puerto Rico and found that both treatments were robust in either format, “...suggesting that the group may be a cost-effective means of providing efficacious treatment.” Comparing CBT and IPT in a meta-analysis of randomized controlled trials (RCTs), Zhou et al. (2017) finds that differences in treatment efficacy seem to vary according to different outcome measures (such as the Beck Depression Inventory vs. the Hamilton Rating Scale for depression).

StrongMinds uses community workers it recruits and intensively trains to lead IPT-G sessions. However, scaling up such treatment to communities across the country (and elsewhere) and making it sustainable requires building more partnerships and identifying a cost-effective process of recruiting and training mentors. Towards this end, Strong Minds began working with BRAC Uganda to develop a model to treat adolescent females with depression. BRAC Uganda implements the “Empowerment and Livelihood for Adolescents,” or ELA, program, which operates through community clubs that girls can attend five afternoons a week. The clubs are led by female mentors, who are selected from the community and given a low-intensity training and a small stipend by BRAC Uganda. In addition to providing a safe space and an opportunity to socialize with other adolescent females, the clubs offer vocational and life skills to their members (Bandiera et al. 2018a, 2018b). Evaluations of ELA in Uganda and Sierra Leone have shown them to be effective in reducing teen and out-of-wedlock pregnancies, child marriages, and non-consensual sexual activity. BRAC Uganda is interested in improving the psychological wellbeing of adolescent females in communities served by ELA clubs – both because of the immediate intrinsic value of reducing anxiety and depression, but also because of potential complementarities between improved mental health and human capital accumulation. Our study aims to test whether a version of StrongMinds’ IPT-G, in which ELA club mentors are trained and supported by StrongMinds’ staff to lead 14-week “talk therapy” sessions for depressed adolescent females from the community, can be effective in causing sustained reductions in depression.

Our study aims to address three research questions. First, we want to test whether IPT-G can substantially reduce depression in adolescent females – at least with medium-run sustained effects, and whether this has knock-on effects on human capital accumulation. Interventions that are specifically designed to improve mental health may be more effective in reducing depression in the medium- and longer-run than anti-poverty programs (Lund et al. 2011).⁴ This trial hopes to provide proof of concept for a low-cost and scalable IPT-G to sustainably reduce depression among adolescent females in sub-Saharan Africa, specifically Uganda.

Second, while cash transfers alone may have transitory effects on psychological wellbeing, there is reason to think that they may enhance the effects of a proven therapy intervention, such as CBT or IPT, if offered immediately following therapy. The idea is that, by providing relief from the day-to-day worries of budget constraints and some insurance against negative shocks that are common in LMICs, cash transfers may create the mental space needed to internalize the lessons and practice the behavioral skills obtained during therapy. For example, Blattman, Jamison, and

⁴ Cash transfers have shown some promise in improving mental health in the short-run (Haushofer and Shapiro 2016), but such gains may dissipate quickly after the cessation of transfers (Baird, de Hoop, and Özler 2013) and/or in the longer-run (Baird, McIntosh, and Özler 2019; Haushofer and Shapiro 2018).

Sheridan (2017) assess the effects of providing CBT, cash, or both to ex-combatants in Liberia. They report impacts on an index of positive self-regard and depression/distress 2-5 weeks and 12-13 months after the end of the intervention. The effect of CBT alone was around 0.1 SD and not statistically significant in the short-run and even smaller in the longer run. Similarly, cash alone had no effect on these outcomes at either follow-up. However, CBT combined with cash had a 0.34 and 0.21 SD effect on this mental health index in the short- and longer-term follow-ups, respectively. The combined treatment also improved quality of social networks (peer and family) in the short-term, which dissipated in the longer-run. The authors suggest that receiving cash might have been akin to an extension of therapy, or follow-up therapy, which has been shown to improve outcomes over the longer-run. Our study provides another setting to test the hypothesis that small lump-sum cash transfers attached to psychotherapy can lead to larger reductions in depression, and thus larger impacts on human capital accumulation, in the longer-run than therapy alone.⁵

Third, and finally, our trial aims to test the hypothesis that reducing depression and dysfunction during adolescence can translate into improved human capital accumulation. In the short-run our primary and secondary outcomes focus on changes in aspirations, beliefs and expectations as mechanisms for future human capital investment. These mechanisms include changes in desired fertility, time preferences, life expectancy, expectations of future paid employment, educational aspirations, and ideal age of first pregnancy and marriage. In the medium-term, we look at actual changes in human capital including school enrollment and skill-based competencies, incidences of pregnancy and marriage, risky sexual behavior, and self-efficacy.

2. Research Design

2.1. Study Setting

Following independence in 1964, there has been significant progress recently on key measures of welfare in Uganda. Poverty headcount ratio at \$1.90/day has been reduced from 63% in 1996 to 42% in 2016, whereas life expectancy at birth has from 44 in 1996 to 60 in 2017 (World Bank, 2018). An officially approved mental health policy does not exist in Uganda. While it is not possible to determine actual mental health financing and expenditures, as it is funded as an integrated component of primary health care, financing has gone up significantly within the last decade because of an ongoing project funded by African Development Bank for strengthening mental health service delivery (World Health Organization 2011). The majority of primary health

⁵ In a trial examining the potential of adding a cash transfer component to a gender transformative mentoring intervention for females in early adolescence in Liberia, called *Girl Empower*, Özler et al. (2020) finds that the program, with or without cash incentives for participation, demonstrated substantial impacts on gender attitudes, life skills, and sexual and reproductive health behaviors. Evidence suggests that program impacts on child marriage, teen pregnancy, and risky sexual behavior were substantially enhanced by adding small cash transfers to caregivers, conditional on girls' program participation.

care doctors have received official in-service training on mental health since 2006, but the majority of primary health care nurses have not. Official referral procedures from primary care to secondary/tertiary care do exist.

The number of mental health professionals is still relatively small, especially the psychosocial staff. Fewer health workers choose to specialize in mental health and a relatively large number of mental health professionals are employed to work as general health workers. An assessment of Uganda's mental health system in 2005 depicted Uganda's Mental Health Act, passed in 1964, as being outdated, with the distribution of mental health care services, personnel and supplies across the country largely inadequate and favoring urban areas (Cooper et al. 2010).

2.2. Interventions

The interventions are implemented by BRAC Uganda, in collaboration with StrongMinds Uganda (SMU). The study includes 106 villages surrounding 106 ELA clubs. BRAC Uganda has established ELA branch offices, each of which oversees a number of clubs, and eight branches (six urban and two semi-urban) were chosen for this study.

After baseline data collection, but prior to the start of the intervention, block randomization (using Stata 14) was used to assign the villages surrounding the ELA clubs into each of the study arms. Randomization was stratified by: (a) BRAC branch office (which includes between 8 and 15 clubs) and (b) above/below median severity of depression among screened-in adolescents at the community level as measured by the eight-item Patient Health Questionnaire depression scale (PHQ-8) to ensure baseline balance on one of the primary outcomes of interest. Treatment is assigned at the cluster (club) level, minimizing the threat of contamination across different treatment arms.

Blinding of intervention units, i.e. ELA clubs, to their own treatment (or that of others) is not feasible. However, the clubs are sufficiently distant from each other so as to be unaware of treatments taking place at other clubs in the same branch or other branches. Even if some club members became aware of the IPT-G intervention nearby, they would not have access to it. The lump-sum cash transfers offered in the third arm will be announced after the rapid resurveys at the end of the IPT-G intervention and, hence, unbeknownst to both the would-be recipients and members at other clubs until then. The three study arms are as follows:

- i. **Control group (36 clubs):** ELA clubs to operate as usual with no other intervention present until program close.
- ii. **IPT-G (35 clubs):** ELA clubs to operate as usual, with a 14-week IPT-G intervention delivered by ELA club peer mentors trained and supported by SMU staff.

- iii. **IPT-G+ (35 clubs):** Same as the IPT-G intervention arm above, followed by a one-time lump-sum cash transfer in the amount of 250,000 Ugandan Shillings (USD 69 at nominal exchange rates as of September 6, 2019), announced and delivered at the end of therapy.

StrongMinds Uganda (SMU) is focused on treating depression in Uganda by training community members to act as facilitators of group-based IPT, providing technical assistance to facilitators and groups, and conducting ongoing monitoring and evaluation of these community-based efforts. For this study, SMU trained ELA mentors⁶ in the treatment communities to serve as IPT-G facilitators, with SMU staff providing supervision and support to these “lay therapists.”⁷

Cash transfers will be delivered by BRAC Uganda immediately following the end of the IPT-G in the third study arm. The transfers are unconditional, in the sense that all study participants who were screened into the study in clubs in the IPT-G+ treatment arm will be offered the transfers. The framing of the transfers is that they are a part of the intervention that is aiming to improve the wellbeing of the subjects.⁸

The readers will have noted that the study design deviates from a classical 2x2 factorial design because it does not have an arm that is ‘cash only.’ While a 2x2 factorial design would have been desirable, BRAC Uganda was unwilling to implement a cash-only arm – mainly because, unlike mental health support and services, cash transfers are not something they are currently considering to add to their adolescent programming.⁹ We discuss the implications of the limitations to our study due to the lack of a ‘cash-only’ arm in Section 4 (Interpreting Results).

2.3. Basic methodological framework/identification strategy

The trial is a cluster-RCT, which is most suitable for the group-based interventions targeted at adolescent females at hand. Individual randomization is undesirable due to the possibility of spillover effects, which would not only pose a threat to identification, but also raise ethical questions. Randomization was considered a suitable methodology for three main reasons. The first is equipoise: it is unknown whether IPT-G model adopted by SMU for adult women in Uganda will improve outcomes for adolescents – by itself or in combination with a cash transfer, particularly when implemented through BRAC ELA clubs and mentors. Second, current funding

⁶ Please see Appendix A for a description of the criteria to be a mentor.

⁷ Please see Appendix B for a more detailed discussion of IPT-G and Appendix C for StrongMinds Uganda’s adaptation and delivery of IPT-G for this study.

⁸ The transfers will be delivered via mobile money directly from BRAC’s main office to the girl’s accounts, or in person if the adolescent does not have a mobile money account. ELA mentors will not be involved or associated with the cash transfers. At the conclusion of the rapid resurvey, the enumerator will inform the girls that they will be receiving a one-time lump-sum cash transfer, collect appropriate consent, and collect their mobile money account information.

⁹ BRAC is not alone among implementers of programs for adolescents in its reluctance to consider cash transfers alone in their programming. A recent study evaluating *Girl Empower*, a mentoring program for early adolescent females implemented by the International Rescue Committee, had a similar design (control, mentoring, and mentoring plus cash) excluding a cash-only arm.

constraints, without a proof of concept, limit the ability to treat all BRAC ELA clubs. Third, the evidence generated through this RCT will inform any potential future scale-up within BRAC and hopefully design of similar mental health interventions for adolescents in other setting.

2.4. Objective and Specific Aims

Our study aims to address three research questions. *First*, we want to test whether IPT-G can substantially reduce depression in adolescent females – at least with medium-run sustained effects. This trial hopes to provide proof of concept for a low-cost and scalable IPT-G to sustainably reduce depression among adolescent females in sub-Saharan Africa, specifically Uganda. *Second*, our trial will test the hypothesis that cash transfers may enhance the effects of a proven therapy intervention, such as CBT or IPT. Following Blattman, Jamison, and Sheridan (2017), which assessed the effects of providing CBT, cash, or both to ex-combatants in Liberia, our study provides another setting to test the hypothesis that small lump-sum cash transfers attached to psychotherapy can lead to larger reductions in depression in the longer-run than therapy alone.¹⁰ *Third*, and finally, our trial aims to test the hypothesis that reducing depression during adolescence can translate into improved human capital accumulation.

In terms of mental health and well-being (see Table 1a), the primary outcomes of the study are measured at baseline, at the end of the 14-week IPT-G intervention, i.e. approximately six months after baseline, and at 12 and 24 months after baseline. These outcomes include two binary indicators of psychological wellbeing: (i) not suffering from psychological distress using the 12-item General Health Questionnaire, or GHQ-12 (GHQ-12 <3) (Goldberg and Williams 1988) and (ii) minimal depression (PHQ-8 ≤4) (Kroenke et al. 2009; Razykov et al. 2012). We also have five secondary outcomes: (i) score on the Rosenberg self-esteem scale (0-30) (Rosenberg 1965), (ii) the score on the Child and Youth Resilience Measure-Revised (0-34) (Jefferies et al. 2018), (iii) locus of control (1-10), (iv) discrete GHQ-12, and (v) discrete PHQ-8. We will present impacts on the mean score and show the cumulative distribution functions of these scores by treatment arm.

¹⁰ Please see “Section 4. Interpreting Results” for a discussion of the lack of a “cash only” arm and its implications for the interpretation of our findings when comparing the medium-run effects of IPT-G and IPT-G+.

Table 1a: Primary and Secondary Outcomes (Mental Health and Well-Being)

Measure	Definition	Baseline	Rapid Resurvey	12-Month	24-Month
=1 if Minimal Depression	Primary Health Questionnaire-8 (PHQ-8) less than or equal to 4	Yes	Primary	Primary	Primary
=1 if does not suffer from pschological distress	General Health Questionnaire-12 (GHQ-12) less than 3	Yes	Primary	Primary	Primary
Score on PHQ-8	Continuous score: 0-24	Yes	Secondary	Secondary	Secondary
Score on GHQ-12	Continuous score: 0-12	Yes	Secondary	Secondary	Secondary
Self-Esteem	Score on the Rosenberg Self-Esteem Scale (0-30)	Yes	Secondary	Secondary	Secondary
Resilience	Score on the Child and Youth Resilience Measure-Revised (0-34)	Yes	Secondary	Secondary	Secondary
Locus of control	On a scale of 1-10, how much control do they feel they have over their lives	Yes	Secondary	Secondary	Secondary

As these scales measure different concepts, we will not construct an index of psychological wellbeing that would lack meaning and would be hard to interpret if there is heterogeneity in program impacts. PHQ-8 and GHQ-12 are both depression scales, but the reason we employ both as outcome measures is that the PHQ-8 is also used by IPT-G facilitators as a monitoring and evaluation tool during the therapy period (administered three times during the 14-week therapy intervention). As the repeated use of the PHQ-8 might differentially affect the responses to this scale at least at the rapid resurvey, we are employing the GHQ-12, which employs different questions, wording, and Likert scales. We are less worried about differential reporting due to this issue at the 12- and 24-month follow-ups, but will employ both primary outcomes (and all five secondary ones) for consistency throughout the study.

We also propose a set of primary and secondary outcomes capturing human capital accumulation (see Table 1b).

Table 1b: Primary and Secondary Outcomes (Human Capital Accumulation)

Measure	Definition	Baseline	Rapid Resurvey	12-Month	24-Month
Desired Fertility	Desired number of children at time of survey	No	Primary	Secondary	Secondary
Time preferences	=1 if prefers 110,000 Uganda Shs one month from now vs. 90,000 today	No	Primary	Secondary	Secondary
Expectations of paid work	Probability (0-10) of engaging in paid work at age 25	No	Primary	Secondary	Secondary
Skill-based Competencies	Total score (0-3) on set of questions on ability to make change that test numeracy and literacy	Yes	Primary	Primary	Primary
Life Expectancy	Number (0-10) of 10 young women just like respondent alive at 40	No	Secondary		
Education Aspirations	Highest desired qualification	No	Secondary		
Desired age of first pregnancy	Number of years in which they hope to first get pregnant (0 if already pregnant)	Yes	Secondary		
Desired age of first marriage	Number of years in which they hope to get married (0 if already married)	Yes	Secondary		
Risky Sex	=1 if no condom use at last sex	Yes		Primary	Primary
Incidence of marriage	Change in ever married	Yes		Primary	Primary
Incidence of pregnancy	Change in ever pregnant	Yes		Primary	Primary
School enrollment	=1 if Enrolled in School	Yes		Primary	Primary
Self-Efficacy	Total score on Schwarzer and Jerusalem 1995 Self-Efficacy Scale (10-40)	No		Primary	Primary

Note: Except self-efficacy at the rapid resurvey, all outcomes will be collected at all rounds regardless of whether they were specified as a (primary or secondary) outcome or not.

At the rapid resurvey, primary and secondary outcomes focus on the expectations of these adolescent females and young women with respect to their life expectancy; desired age at first marriage, first pregnancy, and total fertility; aspirations for schooling and labor force participation; along with their time preferences. These data, combined with evidence of substantial improvements in psychological wellbeing in the treatment groups, would indicate whether sustained gains in human capital accumulation is reasonable to expect in the future and, if so, through what channels. At the 12- and 24-month follow-ups, we propose to measure primary and secondary outcomes that are further down the causal pathway, i.e. potential knock-on effects of improved mental health on human capital accumulation. They include an indicator for being enrolled in school; the score on a set of competencies (skills) using a module developed by the investigators; risky sexual behavior (condom use during last sexual encounter); the incidences of marriage and childbearing since baseline; and self-efficacy.

2.5. Sampling

2.5.1. Sampling strategy

As mentioned above, the study will be conducted in the six urban and two semi-urban branches located in Kampala, all of which are located in the central region. In each branch, 8 to 15 ELA clubs¹¹ were identified with an average of 13 ELA clubs per branch and a total of 106 out of 115 ELA clubs selected for the study.¹² Eligible individuals were adolescent girls aged 13 to 19 years old who resided within a 500-meter radius of an ELA club, and scored 10 or higher on the PHQ-8 during the listing exercise. Scoring at or above the cutoff score of 10 indicates a risk of moderate or severe depression.

In order to identify eligible adolescent females – whether attending ELA clubs or not – the data collection teams established catchment areas within a 0.5-km radius around selected ELA club meeting locations. The team supervisors, with the help of the ELA mentors, identified up to four landmarks to construct a perimeter (the boundaries) of the catchment area, and divided the area into quarters which were assigned to individual team members for baseline data collection.

Team members then administered the listing questionnaire to every single household in their quarter. The instrument, which was digitally programmed on Survey CTO and administered using tablets, was administered to adult household members and included questions aiming to (i) identify all female household members aged 13-19; (ii) obtain their contact information; and (iii) learn about their availability during the next few days.

Additionally, if the young women and/or her caregivers happened to be available at the time, the questionnaire was programmed to (i) obtain consent and/or assent as appropriate; and (ii) administer the PHQ-8 to determine the eligibility of consenting/assenting individuals. If the target individual was not present during the listing visit or the target individual was a minor and no adults were present to provide consent, then the enumerator teams returned at a more convenient time when the adolescent female and/or one of her caregivers/guardians was present to conduct the screening followed by baseline interviews for those who were screened in. No baseline surveys were conducted during the initial listing visit, regardless of the availability of the eligible target individual at the time. The listing exercise produced an average of 19 adolescent girls per community, who were recruited into the study (i.e. aged 13-19, scored 10 or higher on the PHQ-8, and provided assent and/or consent to participate in the study), resulting in a sample 1919 individuals in 106 study communities.¹³

¹¹ While ELA clubs still exist with mentors that can lead therapy, less than 5% of the 13-19 year old adolescents in these communities attended ELA clubs at baseline thus limiting concerns of spillovers within these clubs, particularly in the IPT-G+ arm.

¹² The nine excluded clubs were found no longer to exist or function regularly.

¹³ We have basic information on the universe of 13-19-year-old females in the catchment area of 106 ELA clubs, eligible or not, from the short questionnaire that was administered during listing. However, the study sample only includes those who were deemed eligible as described above.

2.5.2. Statistical power

One of our primary outcome indicators for psychological wellbeing is minimal depression, which is based on the PHQ-8 indicator. It is scored just like the PHQ-9 and has the same cutoff points for severity of depression, meaning that our primary outcome indicator takes on a value of 1 if the PHQ-8 score is between 0 and 4, and 0 otherwise.¹⁴

In 2017, working with our implementing partners, BRAC Uganda and SMU, we conducted a small pilot in 10 ELA clubs belonging to two ELA branches. Trialing a screening and sampling strategy that is similar to the one employed for this study and described above, we screened in 246 females from the catchment areas of these 10 clubs, non-randomly assigned each individual to receive IPT-G or not, conducted IPT-G for 12 weeks and then measured PHQ-9 scores again.¹⁵ These pilot data give us a chance to estimate the relationship between the baseline PHQ-9 score and the follow-up prevalence of minimal depression – with the following important caveats:

- i. The allocation to IPT-G or not was not random, it was decided by SMU based on various baseline criteria;
- ii. The sample size, 204 after accounting for loss to follow-up, was small, and;
- iii. There was significant non-compliance with the treatment assignment (17% assigned to treatment never attended a single session, while 29 out of 63 assigned to control, or 46%, attended at least one IPT-G session).¹⁶

As expected, the data from the pilot phase show that baseline PHQ-9 scores are negatively correlated with minimal depression at follow-up: each unit increase in the PHQ-9 score above the cutoff score of 10 at baseline is associated with a 2.5 percentage point (pp) decrease in the likelihood of minimal depression at follow-up. PHQ-9 scores in the control group were substantially lower at the 16-week follow-up than they were at baseline (6.6 vs. 11.6), with a minimal depression (PHQ-9≤4) rate of 41.3% at follow-up compared with zero (by definition of eligibility) at baseline. Noting that everyone screened into the pilot study had a PHQ-9 score of 10 or higher at baseline, this finding is “...consistent with the literature showing that major

¹⁴ PHQ-9 (and PHQ-8) defines categories of depression severity as follows: none-minimal (0-4); mild (5-9); moderate (10-14); moderately severe (15-19); and severe (>20) (Kroenke et al. 2002). Our study screened in individuals categorized as being at risk of moderate or higher depression severity.

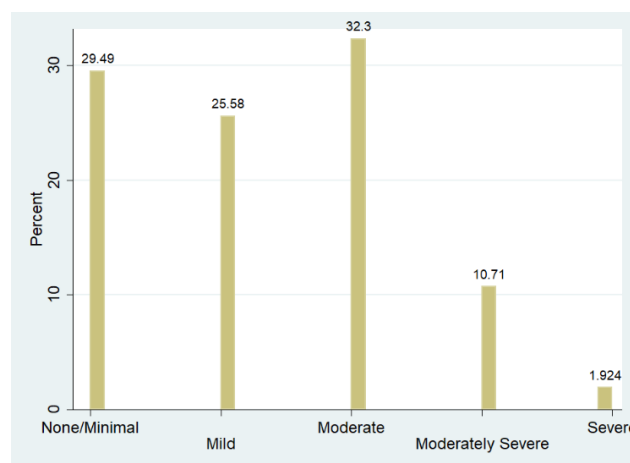
¹⁵ In the pilot phase we used PHQ-9, which was the standard tool being used by SMU prior to our study. Discussions with SMU since then has resulted in the study team choosing to utilize PHQ-8, because the research team did not feel that the enumerators were conformable or trained appropriately to administer question #9 on suicidal ideation. Given that the PHQ-8 is considered equally valid, this was deemed a lower risk strategy.

¹⁶ The uptake of the IPT-G among those screened in and offered to participate was high at 83%. We would like to emphasize that the high non-compliance rate of 46% was a function of the individual assignment to IPT-G within the same clubs. The main study, in contrast to the pilot intervention described above, is a cluster-RCT and will make it practically impossible for any individuals from the control group to receive IPT-G during the study period. Assuming SMU continues to remain active in the region, they will offer IPT-G to the same target population after the completion of this study, i.e. the 24-month follow-up.

depression episodes subside over time even without treatment” (Baranov et al. 2018). That the resolution rate of moderate and severe depression without treatment is similar to other studies in the literature increases our confidence in the data from the pilot phase and suggests that they have some validity. The minimal depression rate was 14.4 percentage points (or 35%) higher in the group assigned to IPT-G compared to the control, suggesting that the IPT-G alone might be highly effective in improving the resolution of depression symptoms.¹⁷ Given the imperfections of the pilot study caveated above and an improved intervention using the lessons from the pilot phase, it is reasonable to expect similarly-sized or larger treatment effects in the main study.

Moving to the baseline data we collected between May and August 2019, we observe that the prevalence of minimal depression is 29.5 percent among all individuals who were screened using the PHQ-8 (into or out of our study) in the study communities during the listing exercise. Figure 1 below shows the full distribution.

Figure 1: Distribution of Depression (based on PHQ-8) in Full Baseline Sample (ineligible and eligible)



If being screened into our study is orthogonal to having minimal to no depression at follow-up, then we might expect the mean of our primary outcome indicator in the control group to be the same, i.e. 0.30. If, however, individuals who were screened into our study are less likely to be minimally depressed at follow-up than those who were screened out, then we might observe prevalence rates of minimal depression lower than 0.30 in the control group at follow-up.

The baseline data discussed above were collected in 106 clusters (catchment areas of BRAC ELA clubs), giving us 35 clusters per treatment arm (and 36 in the control). The average number of eligible females screened into our study in these 106 clusters is 19. Table 2 below indicates the prevalence of *minimal depression* in each treatment arm that would allow us to detect statistically

¹⁷ Baranov et al. (2018) report that the postpartum depression rate among prenatally depressed women was 52% at the six-month follow-up, indicating a spontaneous resolution of the symptoms for approximately half of the women. In their setting, cognitive behavioral therapy further reduced this rate by more than 50% among this group.

significant differences (compared with the control group and with each other) at the 95% level of confidence with 80% power. We report the implied minimum detectable effect (*MDE*) sizes in parentheses and note that:

- i. The ***intraclass correlation of 0.05*** comes from our baseline data with each ELA club as a cluster (it is 0.038 for the discrete PHQ-8 scale and 0.047 for the primary outcome indicator for minimal depression, which have been rounded to 0.05 for both indicators)¹⁸;
- ii. The ***coefficient of variation in cluster size (0.3)*** also comes from our baseline data, using the number of screened-in subjects per cluster;
- iii. To account for any power gains from including the baseline value of the PHQ-8 as a covariate adjustment in the regression models estimating treatment effects at follow-up, we vary the expected ***correlation coefficient*** between the baseline PHQ-8 score among those screened into our study and the probability of minimal depression at follow-up to be low (0.1), medium (0.5), or high (0.9).¹⁹

Table 2: Minimum detectable effects for minimal depression (*proportion*)

Mean of binary indicator for minimal depression at follow-up in control	Mean in IPT-G or IPT-G+ vs. control (<i>MDE</i>)	Mean in IPT-G+ vs. IPT-G (<i>MDE</i>)
0.10 ($\rho = 0.9$)	0.13 (0.03 or 0.10 SD)	0.16 (0.03 or 0.10 SD)
0.20 ($\rho = 0.5$)	0.28 (0.08 or 0.20 SD)	0.37 (0.09 or 0.22 SD)
0.30 ($\rho = 0.1$)	0.40 (0.10 or 0.22 SD)	0.51 (0.11 or 0.24 SD)

ICC=0.05, k=35 (per arm), m=19 (per cluster), alpha=0.05, beta=0.8, coefficient of variation of cluster sizes=0.3, correlation between baseline PHQ-8 and minimal depression (ρ)=0.1, 0.5, or 0.9. MDEs in column 3 are calculated assuming an effect size for IPT-G that is equal to the MDE presented in column 2.

Column 2 shows the MDEs in either treatment when compared with the control group. The MDEs range from 3 to 10 percentage points (pp), depending on prevalence of minimal depression and the correlation coefficient between baseline covariates and minimal depression at follow-up in the control group. These MDEs are lower than the effect observed during the pilot phase (14 pp) and also in line with or lower than estimates from the literature (please see the brief discussion of impact estimates from the literature later in this sub-section).

Column 3 shows the MDE in IPT-G+ (which is IPT-G plus cash) when compared with the IPT-G only group. In our calculations, we assume that the mean at follow-up in the IPT-G group will be equal to the figure in column 2, which is equal to the mean in control plus the MDE. Then, we calculate what the MDE for IPT-G+ would be – over and above the impact of IPT-G

¹⁸ The ICC is 0.038 for PHQ-8 in our baseline data ***among the eligible study sample***. However, by definition, minimal depression is equal to zero for everyone in the study sample. Therefore, we have no choice but to estimate ICC for minimal depression using the universe of screened individuals – eligible or not.

¹⁹ We note that we consider low autocorrelation, i.e. the conservative estimates with higher MDEs, to be more plausible based on data from the pilot phase of this study and our reading of the extant literature.

alone. These MDEs are between 3-11 pp, which imply impact estimates of 6-21 pp for IPT-G+ in comparison with the control group.

Table 3 presents the same calculations for PHQ-8, the discrete scale that underlies the binary indicator of minimal depression. The mean (SD) of PHQ-8 in our baseline data in the target population is 8.14 (5.18). Again, the last row assumes one extreme of little correlation between being screened into our study and follow-up scores and uses the mean (SD) PHQ-8 score in the entire target population. At the other extreme, the first row considers the case of almost perfectly correlated baseline and follow-up PHQ-8 scores and uses the mean (SD) PHQ-8 score, 13.08 (2.88) among those screened into our study, i.e. among those with PHQ-8 ≥ 10 . Hence, these two rows provide reasonable bounds for the mean PHQ-8 score at follow-up in the study sample.

Table 3: Minimum detectable effects for the PHQ-8 scale (mean)

Mean (SD) of PHQ-8 at follow-up in control	Mean in IPT-G or IPT-G+ vs. control (MDE)	Mean in IPT-G+ vs. IPT-G (MDE)
13.08 (2.88) ($\rho = 0.9$)	12.80 (0.28 or 0.1 SD)	12.52 (0.28 or 0.1 SD)
8.14 (5.18) ($\rho = 0.1$)	7.01 (1.13 or 0.22 SD)	5.88 (1.13 or 0.22 SD)

ICC=0.05, k=35 (per arm), m=19 (per cluster), alpha=0.05, beta=0.8, coefficient of variation of cluster sizes=0.3, correlation between baseline PHQ-8 and minimal depression (ρ)=0.1 or 0.9. MDEs in column 3 are calculated assuming an effect size for IPT-G that is equal to the MDE presented in column 2.

As can be seen, the MDEs in either treatment arm compared with the control group are small at either extreme, ranging from 0.28 when autocorrelation is high to 1.13 when it is low. These MDEs represent standardized effect sizes of 0.1 and 0.22, respectively, with the detectable effect ranging from a reduction of 2% (row 1) to 14% (row 2) of the mean in the control group. Again, these MDEs are smaller than the suggested effect sizes during the pilot phase.²⁰ We note that we consider the conservative estimates, i.e. the higher MDEs, to be more plausible based on data from the pilot phase of this study and our reading of the extant literature.

These MDEs seem adequate for our study considering our estimates of the impact of IPT-G on minimal depression and PHQ-9 from the pilot phase. Again, with the caveats made above regarding the pilot phase of our study, our best guess of the effect of IPT-G is 12-14 pp on minimal depression (compared with the 7-10 pp MDEs presented above) and 1.3 to 1.9 on the PHQ-8 score (compared with the 0.67-1.21 MDEs presented above).²¹ Please note that our effect estimates

²⁰ At the rapid resurvey immediately following the IPT-G intervention (approximately 5-6 months after baseline), we will combine the two treatment arms (IPT-G and IPT-G+) as these surveys take place before the announcement of lump-sum cash transfers in the IPT-G+ group. The standardized MDE for PHQ-8 declines from 0.22 in either treatment arm to 0.18 in the combined treatment group. Similarly, we can combine the 12- and the 24-month follow-up rounds to estimate an average treatment effect over a period of 12 months between the 12- and 24-month follow-ups. Simulations suggest that the standardized MDE for PHQ-8 would decline from 0.22 to 0.15 in the case of very low autocorrelation ($\rho < 0.1$), increasing towards 0.22 with higher values of ρ towards 1.

²¹ We did not implement IPT-G plus cash during the pilot phase, focusing our energies on getting the IPT-G intervention right. In light of lessons from the pilot phase, SMU decided to increase the length of the therapy intervention from 12 weeks to 14.

from the pilot phase are intent-to-treat effects, with almost half of those assigned to the “control” group having attended at least one IPT-G session. With a cleaner cluster-randomized trial and efforts to increase the efficacy of IPT-G by improving the selection and training of mentors for the main study, it is reasonable to expect higher effect sizes during the main study than the pilot.

While the preliminary (and non-experimental) impact estimates from the pilot phase compare well with the MDEs from the main study, it is useful to assess the MDEs in light of other studies that have reported treatment effects on mental health outcomes, preferably among adolescent females and/or for interventions that include therapy or cash transfers. Baranov et al. (2018), examines the effects of offering cognitive behavioral therapy (CBT) to prenatally depressed women in Pakistan on the likelihood of postpartum depression at six and 12 months after treatment, as well as their mental health, economic decision making, financial empowerment, and parental investments seven years later. The intervention reduced the likelihood of a diagnosis of depression by 35 and 39 pp at the 6- and 12-month follow-ups, respectively (over means of 52% and 58% in the control group), although the effect on this particular outcome dissipated by the seven-year follow-up. However, large and statistically significant beneficial effects were sustained on depression severity (0.60, 0.66, and 0.18 SD reductions at 6-, 12-, and 84-month follow-ups, respectively), the likelihood of recovering permanently (23 pp increase over a mean of 34% in the control group), and never recovering from depression (12 pp decrease over a mean of 23% in the control group). Particularly promising and relevant for our study are the knock-on effects of these improvements in mental health on other outcomes, such as future economic decision making, empowerment, and parental investments (especially on female children). This is because our study focuses on younger adolescent females with the hypothesis that improvements in mental health during adolescence might lead to higher human capital accumulation in the medium-run and empowerment in the longer-run.

Blattman, Jamison, and Sheridan (2017) assess the effects of providing CBT, cash, or both to ex-combatants in Liberia. They report impacts on an index of positive self-regard and depression/distress 2-5 weeks and 12-13 months after completion of the intervention. The effect of CBT alone was around 0.1 SD and not statistically significant in the short-run and even smaller in the longer run. However, CBT combined with cash had a 0.34 and 0.21 SD effect on this mental health index in the short- and longer-term follow-ups, respectively. The combined treatment also improved quality of social networks (peer and family) in the short-term, which dissipated in the longer-run. The combination of CBT and lump-sum cash transfers offered at the end of the intervention make this study relevant to ours. However, it should also be noted that the post-conflict setting and the target population of ex-combatants and ex-criminals are somewhat unique

and may have limited relevance for our study, which is targeting younger adolescent females at risk of mild to severe depression.²²

Closer to our setting of group-based interpersonal psychotherapy (IPT-G) in Uganda, Bolton et al. (2003) report the effects of IPT-G on men and women in 30 villages in rural Uganda who met the criteria for major or subsyndromal depression at baseline. Using a cluster-randomized trial, they assessed the impact of a 16-week IPT-G intervention (90-minute sessions per week) on the rate of depression, as well as the severity of depression and dysfunction. The intervention reduced the risk of major depression at follow-up by 26 pp, with the odds of major depression in the control group being 4-5 times higher than the intervention group. Severity of depression and dysfunction were also substantially reduced in the intervention arm, with the ability to perform individual tasks at follow-up being significantly improved, especially among female IPT-G beneficiaries. Comparing individual and group formats of CBT versus IPT, Rosselló, Bernal, and Rivera-Medina (2008) find that CBT and IPT are robust treatments of depression among adolescents in both formats. However, they also find that CBT produced significantly greater decreases in depressive symptoms and improved self-concept than IPT.

Cash transfers alone can have effects on anxiety and depression, especially in the short-term. Baird, de Hoop, and Özler (2013) presents the effects of an unconditional cash transfer (UCT) program on psychological distress (measured by GHQ-12, the 12-item General Health Questionnaire) for a target group of initially never-married females aged 13-22 in Malawi. The program provided \$10 per month over the course of two school years, with an average of \$3/month being transferred directly to the adolescent female with the rest of the money transferred to her parents/guardians. At the 12-month follow-up, UCT beneficiaries were 14 pp (or 38%) less likely to be suffering from a binary indicator of psychological distress and had 20-35% lower GHQ-12 scores compared to the control. These reductions in psychological distress are smaller than the effect of migration from Tonga to New Zealand (Stillman, McKenzie, and Gibson 2009), but similar to the mental health effects among female youths of moving to richer neighborhoods in the U.S. (Kling, Liebman, and Katz 2007). However, these improvements in mental health quickly disappeared once the cash transfer program ended. Lump-sum cash transfers in Kenya did not have effects on cortisol levels or subjective wellbeing approximately one year after the transfers when the transfer levels were around USD 350-400 PPP, but had some beneficial

²² In a trial examining the potential of adding a cash transfer component to a gender transformative mentoring intervention for females in early adolescence in Liberia, called *Girl Empower*, effect sizes on child marriage, onset of sexual debut, and risky sexual behaviors were 50 to 100% larger in the intervention arm that combined mentoring with small cash transfers tied to the girls' program participation than in the arm with mentoring intervention alone (Özler et al. 2019).

effects on depression, stress, life satisfaction, and optimism when the transfers were large (about USD 700 PPP) (Haushofer and Shapiro 2016; Haushofer et al. 2019).

Minimum detectable effects on other human capital accumulation outcomes in the longer-run

While improvements in psychological wellbeing are intrinsically important and are, therefore, primary outcomes both in the short- and the longer-run, they are also the primary mechanism by which human capital accumulation might increase in the treatment arms in the longer-run. Hence, in Table 4 below, we present MDEs for some of our primary outcomes at the 24-month follow-up – namely incidences of being marriage and pregnancy since baseline, share of enrolment in school, and the competencies score (a test of math and reading skills that are important in day-to-day life, such as for market transactions, designed by the authors).

Table 4: Minimum detectable effects for other outcomes (24-month follow-up)

Outcome/follow-up	Mean at follow-up in C (<i>expected</i>)	Mean in IPT-G or IPT-G+ vs. C (<i>MDE</i>)	Mean in IPT-G+ vs. IPT-G (<i>MDE</i>)
Incidence of marriage	0.25	0.18 (0.07 or 0.16 SD)	0.12 (0.06 or 0.14 SD)
Incidence of pregnancy	0.28	0.20 (0.08 or 0.18 SD)	0.13 (0.07 or 0.16 SD)
Enrolled in school	0.25	0.33 (0.08 or 0.18 SD)	0.42 (0.09 or 0.21 SD)
Competencies [<i>range=0-4</i>]	1.31	1.58 (0.27 or 0.18 SD)	1.83 (0.27 or 0.18 SD)

ICCs between 0.015 and 0.035 depending on the outcome, k=35 (per arm), m=19 (per cluster), alpha=0.05, beta=0.8, coefficient of variation of cluster sizes=0.3. MDEs in column 3 are calculated assuming an effect size for IPT-G that is equal to the MDE presented in column 2.

The prevalence of being ever married (pregnant) at baseline is equal to 0.13 (0.18), respectively, and we estimated the incidence to be 0.25 (0.28) – based on the age distribution of the study sample, marriage and pregnancy patterns by age, and studies of adolescents and young women from other settings in sub-Saharan Africa. The MDEs are 7-8 pp (or 25-30% of the control mean) for either treatment vs. the control group and 6-7 pp between the two intervention groups.

These MDEs seem reasonable to expect in light of other studies of adolescent females and young women from Uganda, as well as Liberia and Malawi. Bandiera et al. (forthcoming) evaluates the impacts of girls’ clubs in Uganda and finds teen pregnancies and non-consensual sex to drop by a third and early marriage/cohabitation fall rapidly as a result of the intervention that provides adolescent girls a safe space to gather, along with training modules on life and vocational skills.²³ Özler et al. (forthcoming), in a study design similar to that considered here, evaluates the impact of a 32-week mentoring program targeted at early adolescent females – alone and when combined with small cash transfers tied to weekly participation. The paper finds large effects on an index of sexual and reproductive health 24 months after baseline (and more than one year after the

²³ The paper also finds effects on self-employment, aspirations for schooling and age at first marriage/pregnancy. It is worth noting that the participation (take-up) rate for the intervention is approximately 20%, which is well below the anticipated rate of take-up of IPT-G in our study among the eligible adolescents, who are offered therapy.

completion of the mentoring program) – with an effect of 0.24 SD for the mentoring program alone and 0.37 for mentoring combined with cash. The combined intervention reduced the incidence of child marriage over a course of 24 months by more than 50% (7.8% vs. 3.7%). Finally, unconditional cash transfers, given to adolescent schoolgirls in Malawi for two years, reduced the incidence of marriage by 46% - from 18% in the control group vs. 9.7% in the treatment group (Baird, McIntosh, and Özler 2011).²⁴ Given the evidence, even the means in column 4 (required to detect differences between the two treatment arms when there is already a significant effect in the IPT-G arm) are within the realm of what might be expected from a combined therapy and cash intervention.

Only 41% of the study sample was enrolled in school at baseline and we estimate this share to drop to approximately 25% by the 24-month follow-up. The MDEs on enrollment are slightly larger. In fact, if the IPT-G has a detectable effect on enrollment (which would be a valuable contribution to the literature as experimental estimates of mental health interventions for children on school participation are rare), the enrollment rate in IPT-G+ at follow-up would have to be greater than 0.41 to be detectable, which is equal to mean enrollment at baseline and unlikely to be achieved. Hence, while our study is well-powered to detect school participation effects for either intervention vs. the status quo, it is less powered to detect differences between IPT-G and IPT-G+ if IPT-G is effective in increasing enrollment rates. It is, however, well-powered to detect improvements in skills (competencies), with the MDE equal to 0.18 SD.

2.5.3. Variations from the intended sample

Take-up among those assigned to treatment was high in the pilot phase and StrongMinds Uganda's had similarly high take-up rates in implementing this intervention in adult populations. However, imperfect take-up is always a possibility in real life effectiveness studies like this one and reports from the early stages of the IPT-G intervention from BRAC suggest that take-up among the adolescent target population here might be lower than SMU's experience with adult female populations in Uganda. As the expected effect sizes are sufficiently larger than the MDEs, we do not anticipate this to cause a major problem with inference for ITT estimates. In case of take-up that is substantially lower than 100%, such as 75% or lower, we will also present estimates of the local average treatment effect (LATE) for each intervention arm, using random assignment to that arm as an instrument, which, with neither treatment available to the control group, is equal to the average on the treated (ATT).

²⁴ CCTs to baseline dropouts reduced the incidence of marriage by 27% (from 57.5% to 41.8%) in the same study.

Non-compliance among those assigned to control was a much bigger problem in the pilot study (close to half of those assigned to the control group attended at least one IPT-G session), but we stipulate that this was due to the fact that the assignment to treatment was done at the individual level during the pilot phase – due to limited resources available for this part of the study. The main study is a cluster-randomized trial, with each cluster, i.e. each BRAC ELA club, sufficiently distant from each other, meaning that we are confident that no one in the control group will have access to therapy. Hence, while there might be lower than 100% uptake of IPT-G in either treatment arm, there should be minimal contamination of the control group.

Subjects being lost to follow-up is a problem that is faced by all studies that longitudinally track study subjects over time. We anticipate that there will be some attrition during the two-year period between baseline and our final follow-up and have strategies to minimize attrition and the risk that it is differential between treatment arms. These include intensive **detailed regular tracking** protocols to interview subjects who moved during the study and **intensive tracking protocols** for a random sub-sample of those who could not be located during regular tracking. Any attrition that obtains in the data will be analyzed for differences between treatment arms in levels and in baseline characteristics (as in Baird, McIntosh, and Özler 2019, Table 2). The robustness of impact estimates will be tested using estimates reweighted to account for attrition (IPW), upper and lower bounds on impact estimates for all primary outcomes (Lee, 2009), as well as adjustments using the techniques of Kling and Liebman (2004) (see Appendix Tables S11-S13 in Baird, McIntosh, and Özler 2019).

2.6. Data collection and processes

Data collection involves the use of several questionnaires, including: a listing survey, a long baseline survey, a shorter and rapid follow-up survey, and two follow up surveys that are more comprehensive to measure the short- and medium-term outcomes of interest.

2.6.1. Baseline

Baseline data collection took place from May to August 2019 and covered 106 ELA clubs in the eight branches in and around Kampala. Enumerators used a listing questionnaire to identify 13-19-year-old adolescents and screened those who assented/consented to participate in the study for mild to severe depression using the PHQ-8 scale. The same data collection team then returned to the village to administer a multi-topic baseline survey instrument to eligible individuals screened into the study. The survey was programmed in SurveyCTO and administered by enumerators in

face-to-face interviews. Overall, 45% of adolescents who took the PHQ-8 screened through, and refusal rates were low at 3%.²⁵

The listing and baseline questionnaires can be found [here](#). They cover most of the primary and secondary outcomes, as well as data on individual and household characteristics that (a) are prognostic of the outcomes of interest and (b) might be important sources of impact heterogeneity. A description of data collection modules (Appendix Table D1) and a balance table, which includes baseline values of the primary and secondary outcomes along with the pre-specified covariate adjustments for the impact regressions (Appendix Table D2), are provided in Appendix D.

2.6.2. Midlines and Endline

The intervention started in September 2019 and will run for 14 weeks. At the end of these 14 weeks a rapid resurvey will be conducted (prior to the distribution of cash), which will collect data only on the primary outcomes. Two additional rounds of data collection will take place one and two years after baseline to both assess whether impacts on the primary outcomes are sustained over time as well as any knock-on effects on the secondary outcomes of interest, such as education, skills, fertility and marriage, and exposure to physical or sexual violence.

2.6.3. Pilot data

Please see Section 2.4.2 for a detailed discussion of the pilot phase of the study, which was used to establish feasibility; provide guidance on realistic effect sizes that the study team used for power calculations; and identify weaknesses in program implementation so that the intervention could be improved before the start of the main study.

3. Empirical Analysis

We take advantage of the randomized allocation of the interventions at the cluster (ELA club) level to estimate causal effects in the two treatment arms. At each follow-up, we will analyze the primary and secondary outcomes at the individual level. To estimate intention-to-treat (ITT) effects of each intervention on individual outcomes, we will employ a regression model of the following form for each round of follow-up data collection:

$$Y_{ij} = \alpha + \gamma^1 T_j^1 + \gamma^2 T_j^2 + \beta X_{ij} + \varepsilon_{ij} \quad (1),$$

where Y_{ij} is an outcome variable for individual i in cluster j , T_j^1 and T_j^2 are binary indicators for cluster-level interventions IPT-G and IPT-G+, while X_{ij} is a vector of baseline characteristics.²⁶

²⁵ In a small number of communities, due to population density within the study area, the teams were not able to complete the screening of all 13-19-year-old females before having to move on to the next community. In those clusters, it is possible that the population that was screened was not a random sub-sample of all households with at least one 13-19-year-old female member. The statistics from our study should therefore be interpreted accordingly.

²⁶ We will also report coefficient estimates for “Any Treatment,” which is a pooled indicator for treatment for those in T1 and T2. This will boost power in the rapid resurvey round, as data in that round are collected before the announcement (and transfer) of cash transfers in the IPT-G+ arm.

The standard errors ε_{ij} , clustered at the cluster level, account for both the design effect of the cluster-level treatment and heteroskedasticity inherent in the regression model.

For each outcome measure, we will estimate two versions of the model in equation (1). In the “unadjusted” regressions, we only include indicators for the 16 strata used to perform block randomization. These strata were formed using eight BRAC branch offices and by forming two groups within each office based on the mean PHQ-8 score at each club (Bruhn and McKenzie 2009). In the “adjusted” regressions, we will include centered indicator variables for age in years, the poverty probability index (PPI) as a measure of wealth²⁷, an indicator for having started childbearing, an indicator for currently married, and the baseline (lagged) value of the PHQ-8 – each of which will be interacted with the treatment indicators.²⁸ These variables were chosen because they are expected to be sources of impact heterogeneity and prognostic of outcomes at follow-up, the latter of which implies that their inclusion might improve the precision of the impact estimates, which was conservatively taken into account in the power calculations presented above. We prefer this analysis of covariance specification to a difference-in-difference estimation because of the expected gains in power (McKenzie 2012).

As mentioned above in Section 2.5.3, in case of take-up rates lower than 75%, we will report the LATE/ATT estimates, using randomized assignment to IPT-G (IPT-G+) as an instrument for receiving therapy (therapy + cash). However, while the comparison of ITT estimates is straightforward between the two treatment arms, a comparison of ATTs is harder to interpret. Furthermore, such a comparison is of limited interest to the research question of whether combining group therapy with cash transfers yields impacts that are larger and/or sustained for a longer period of time. Instead, we can examine the marginal effect of adding cash transfers to group therapy, by conducting the analysis among those who took up therapy in either treatment arm. While those who took up therapy are a selected group, they should not differ between the two treatment arms because the cash transfers were not announced until the completion of the 14-week therapy program. We can confirm balance of baseline characteristics between IPT-G and IPT-G+ within this selected sub-sample at follow-up and proceed to analyze whether the addition of cash transfers caused differential impacts among those (a) who took up therapy and (b) who did not. This secondary analysis is non-experimental and would have lower power (akin to heterogeneity analysis), but would shed

²⁷ Please see <https://www.povertyindex.org/blog/new-ppi-construction-methodology> for details on the construction methodology for the PPI.

²⁸ Per Lin (2013), this is the preferred specification when using regression adjustments in experiments.

some light on the mechanism by which cash transfers may or may not enhance the effects of group therapy among adolescent females at risk of moderate to severe depression at baseline.

3.1. Multiple outcome and multiple hypothesis testing

We will present q-values controlling for false discovery rates (FDR), as described in Anderson (2008). We will use Anderson's Stata code to calculate FDR-adjusted q-values, which uses a simple method proposed by Benjamini and Hochberg (1995) to calculate the smallest level of significance at which the null hypothesis would be rejected. We will conduct multiple hypothesis testing within the two domains. Within mental-health and well-being we have 6 q-values for the two primary outcomes (two outcomes times three pairwise comparisons per outcome) and 15 q-values for the five secondary outcomes; while for human capital accumulation, the total primary and secondary outcomes vary by round. Q-values will be calculated at each follow-up round and presented in separate summary tables of impacts on the primary and secondary outcomes.

3.2. Heterogeneous effects

The adjusted regression specification automatically conducts heterogeneity analysis, as it is fully interacted with baseline covariate adjustments that were selected for this purpose. We will report these in a separate sub-section (rather than along with the main impact estimates), but note that our study is not powered to detect heterogeneity of program impacts with precision.

4. Interpreting results

One of the issues that can cause difficulties in the interpretation of the findings from this study comes from the fact that it does not have a classical 2x2 factorial design that includes a “cash alone” arm. While it would have been ideal, from a study design perspective, to have such a design, which would have enabled us to experimentally reject (or not) that cash alone would have been as effective as IPT-G+. Such benchmarking of the (cost-) effectiveness of a group therapy intervention to cash transfers alone would have been nice, the three-arm trial we designed for our setting, leaving out a “cash-only” arm, is an adequate study design for the following three reasons.

First, the extant evidence on the effects of economic interventions in general, or cash transfers in particular, do not support the idea of improved mental health outcomes past the short-run. For example, in their review, Lund et al. (2011) finds that the mental health effects of poverty alleviation programs were inconclusive. Blattman, Jamison, and Sheridan (2017) find no effects of lump-sum cash transfers alone on mental health outcomes in the short- or the medium-run. Short-term effects of cash transfers (monthly or lump-sum) on psychological wellbeing that were observed in the short-run dissipated a couple of years later (Baird, de Hoop, and Özler 2013; Baird, McIntosh and Özler 2019; Haushofer and Shapiro 2016, 2018). Hence, we do not think that there

is sufficient equipoise to include a “cash only” arm when it comes to sustained effects on depression two-years after the end of the intervention.

Second, even if one could make a case that there may be a sustained income effect on mental health – for example, such as those indicated by studies of lottery winners (Gardner and Oswald 2007; Lindahl 2005) – the amounts offered to IPT-G participants here are too small to have this kind of an effect two years after they are transferred. A similar argument has been made in another experiment that lacked a pure unconditional cash transfer (UCT) arm: Benhassine et al. (2015) argue that the labelling of the cash transfer as an education support program increased school participation through its effect on the “...parents’ belief that education was a worthwhile investment,” rather than through a pure income effect, because the transfers were too small to cause the observed effects.

Finally, and related to the point above, one of the aims of our trial is to test the efficacy of a low-cost and scalable intervention through two NGOs that have a track record of implementing programs that are being utilized here (BRAC Uganda and StrongMinds Uganda). While BRAC Uganda is interested in taking advantage of its ELA girls’ clubs’ platform to provide mental health services across Uganda, it does not have any plans to provide UCTs, especially not in transfer sizes that might perhaps have sustained effects. Hence, the lack of evidence on the potential effectiveness of UCTs on sustained reductions in depression, combined with a lack of interest from the implementing partners, resulted in the study team designing a trial that has only three arms. Should the trial show that IPT-G+ is significantly more effective than IPT-G alone in reducing depression in the medium-run, our interpretation will be that there is a complementarity between the two interventions, and not that cash is effective on its own for sustained improvements in psychological wellbeing.

5. List of References

- Anderson, M.L. 2008. “Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects.” *Journal of the American Statistical Association*, Vol. 103(484): 1481-1495.
- Baird, S., De Hoop, J. & Özler, B. 2013. “Income Shocks and Adolescent Mental Health.” *Journal of Human Resources*, Vol. 48(2): 370-403.
- Baird, S., McIntosh, C., and Özler, B. (2011), “Cash or Condition: Evidence from a Randomized Cash Transfer Program,” *Quarterly Journal of Economics*, Vol. 126(4), 1709-1753.
- Baird, S., McIntosh, C., and Özler, B. (2019) “When the Money Runs Out: Do Cash Transfers Have Sustained Effects?” *Journal of Development Economics*, Vol. 140, 169-185.
- Bandiera, O., Buehren, N., Burgess, R., Goldstein, M., Gulesci, S., Rasul, I. & Sulaiman, M. (2019). “Women’s Empowerment in Action: Evidence from a Randomized Control Trial in Africa.” https://www.dropbox.com/s/f8glsmbnt6s7bfe/PaperComplete_20170416.pdf?dl=0, accessed September 16, 2019.

- Bandeira, O., Buehren, N., Goldstein, M., Rasul, I., & Smurra, A. (2018). "The Economic Lives of Young Women in the Time of Ebola: Lessons from an Empowerment Program." https://www.povertyactionlab.org/sites/default/files/publications/The-Economic-Lives-of-Young-Women_EIA_SL_Bandiera-et-al_Dec2018.pdf, accessed September 16, 2019.
- Baranov, V., Bhalotra, S., Biroli, P., & Maselko, J. (2018). "Maternal Depression, Women's Empowerment, and Parental Investment: Evidence from A Randomized Control Trial." <https://drive.google.com/file/d/1uAMCHmJ7juDa2CKRvHq4K9fwTFJZyw9P/view>, accessed September 16, 2019.
- Benhassine, N., Devoto, F., Duflo, E., Dupas, P. & Pouliquen, V. 2015. "Turning a Shove into a Nudge? A "Labeled Cash Transfer" for Education." *American Economic Journal: Economic Policy*, Vol. 7(3): 86-125.
- Benjamini, Y. & Hochberg, Y. 1995. "Controlling the false discovery rate: a practical and powerful approach to multiple testing." *Journal of the Royal Statistical Society. Series B (Methodological)*, 289-300.
- Blattman, C., Jamison, J.C. & Sheridan, M. 2017. "Reducing Crime and Violence: Experimental Evidence from Cognitive Behavioral Therapy in Liberia." *American Economic Review*, Vol. 107 (4): 1165-1206.
- Bolton, P., Bass, J., Neugebauer, R., and others. 2003. "Group Interpersonal Psychotherapy for Depression in Rural Uganda." *Journal of the American Medical Association*, Vol. 289(23): 3117-24.
- Bruhn, M., and McKenzie., D. 2009. "In Pursuit of Balance: Randomization in Practice in Development Field Experiments." *American Economic Journal: Applied Economics*, Vol. 1(4): 200–32.
- Cooper, S., Ssebunnya, J., Kigozi, F., Lund, C., Flisher, A., & MHaPP Research Programme Consortium. 2010. "Viewing Uganda's mental health system through a human rights lens." *International Review of Psychiatry*, 22(6), 578-588.
- Currie, J. & Stabile, M. 2006. "Child Mental Health and Human Capital Accumulation: The Case of ADHD." *Journal of Health Economics*, Vol. 25(6): 1094–118.
- Dercon, S. & Krishnan, P. 2009. "Poverty and Psycho-Social Competencies of Children: Evidence from the Young Lives Sample in Four Developing Countries." *Children, Youth, and Environments*, Vol. 19(2): 138–63.
- Di Clemente, R. J., Wingood, G.M., Crosby, R.A., Sionean, C., Brown, L.K., Rothbaum, B., Zimand, E., Cobb, B.K., Harrington, K., & Davies, S. 2001. "A Prospective Study of Psychological Distress and Sexual Risk Behavior Among Black Adolescent Females." *Pediatrics*, Vol. 108(5): e85.
- Eisenberg, D., & Golberstein, E. 2009. "Mental Health and Academic Success in College." *The B.E. Journal of Economic Analysis & Policy*, Vol. 9(1): Article 40.
- Evans, G.W., Kim, P., Ting, A.H., Teshler, H.B., & Shanis, D. 2007. "Cumulative Risk, Maternal Responsiveness, and Allostatic Load among Young Adolescents." *Developmental Psychology*, Vol. 43: 341–51.
- Fishbein, D.H., Herman-Stahl, M., Eldreth, D., Paschall, M.J., Hyde, C., Hubal, R., Hubbard, S., Williams, J., and Lalongo, N. 2006. "Mediators of the Stress-Substance-Use Relationship in Urban Male Adolescents." *Prevention Science*, Vol. 7(2): 113–26.
- Fletcher, J. M. 2008. "Adolescent Depression: Diagnosis, Treatment, and Educational Attainment." *Health Economics*, Vol. 17(11): 855–71.
- Fletcher, J.M., and Wolfe, B. 2008. "Child Mental Health and Human Capital Accumulation: The Case of ADHD Revisited." *Journal of Health Economics*, Vol. 27(3):794–800.
- Gardner, J., and Oswald, A.J. 2007. "Money and Mental Wellbeing: A Longitudinal Study of Medium-Sized Lottery Wins." *Journal of Health Economics*, 26(1): 49-60.
- Goldberg, D., & Williams, P. 1988. *A user's guide to the General Health Questionnaire*. Windsor, UK: NFER-Nelson.

- Haushofer, J., & Shapiro, J. 2016. The Short-term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya, *The Quarterly Journal of Economics*, Vol. 131 (4), 1973–2042
- Haushofer, J., & Shapiro, J. 2018. “The Long-term Impact of Unconditional Cash Transfers: Experimental Evidence from Kenya.” http://www.princeton.edu/haushofer/publications/Haushofer_Shapiro_UCT2_2018.pdf, accessed September 16, 2019.
- Haushofer, J., Chemin, M., Jang, C., & Abraham J. 2019. “Economic and Psychological Effects of Health Insurance and Cash Transfers: Evidence from a Randomized Experiment in Kenya.” http://www.princeton.edu/haushofer/publications/Haushofer_Chemin_Jang_Abraham_Insurance_2019-04-24.pdf, accessed September 16, 2019.
- Jefferies, P., McGarrigle, L., & Ungar, M. 2018. The CYRM-R: a Rasch-validated revision of the Child and Youth Resilience Measure. *Journal of Evidence-Informed Social Work*, 1-24.
- Heckman, J.J., Stixrud, J. & Urzua, S. 2006. “The Effects of Cognitive and Noncognitive abilities on Labor Market Outcomes and Social Behavior.” *Journal of Labor Economics*, Vol. 24(3): 411–82.
- Kessler, R.C., Foster, C.L., Saunders, W.B., & Stang, P.E. 1995. “Social Consequences of Psychiatric Disorders, I: Educational Attainment.” *American Journal of Psychiatry*, Vol. 152: 1026–32.
- Kling, J., and J. Liebman. 2004. “Experimental Analysis of Neighborhood Effects on Youth,” *Unpublished manuscript*.
- Kling, J.R., Liebman, J.B., and Katz, L.F. 2007. “Experimental Analysis of Neighborhood Effects.” *Econometrica*, Vol. 75(1):83–119.
- Krishnan, P., & Krutikova, S. 2012. “Noncognitive Skill Formation in Poor Neighborhoods of Urban India.” Cambridge Working Papers in Economics 1010.
- Kroenke, K., & Spitzer, R.L. 2002. “The PHQ-9: a new depression diagnostic and severity measure.” *Psychiatric Annals*, Vol. 32: 509-521.
- Kroenke, K., Strine, T. W., Spitzer, R. L., Williams, J. B. W., Berry, J. T., & Mokdad, A. H. 2009. “The PHQ-8 as a measure of current depression in the general population.” *Journal of Affective Disorders*, Vol. 114(1-3): 163-173.
- Lee, D.S. 2009. Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *The Review of Economic Studies*, Vol. 76(3): 1071-1102.
- Lin, W. 2013. “Agnostic Notes on Regression Adjustments to Experimental Data: Reexamining Freedman’s Critique.” *The Annals of Applied Statistics*, Vol. 7(1): 295-318.
- Lindahl, M. 2005. “Estimating the Effect of Income on Health and Mortality Using Lottery Prizes as an Exogenous Source of Variation in Income.” *The Journal of Human Resources*, 40(1): 144-168.
- Lund, C., de Silva, M., Plagerson, S., Cooper, S., Chisholm, D., Das, J., Knapp, M., & Patel, V. 2011. “Poverty and Mental Disorders: Breaking the Cycle in Low-Income and Middle-Income Countries.” *The Lancet*, Vol. 378: 1502–14.
- McKenzie, D. 2012. “Beyond Baseline and Follow-up: The Case for more T in Experiments.” *Journal of Development Economics*, 99(2): 210-21.
- McLoyd, V.C., Kaplan, R., Purtell, K.M., Bagley, E., Hardaway, C.R., & Smalls, C. 2009. “Poverty and Socioeconomic Disadvantage in Adolescence.” In *Handbook of Adolescent Psychology, 3rd edition*, ed. Richard M. Lerner and Laurence D. Steinberg, 444–91. New York: John Wiley.
- Özler, B., Hallman, K., Guimond M.-F., and Kelvin, E. 2020. “Girl Empower – A Gender Transformative Mentoring and Cash Transfer Intervention to Promote Adolescent Wellbeing: Impact Findings from a Cluster-Randomized Controlled Trial in Liberia.” *Social Science and Medicine – Public Health*, Vol. 10.
- Patel, V., Fisher, A.J., Hetrick, S. & McGory, P. 2007. “Mental Health of Young People: a Global Public Health Challenge.” *The Lancet*, Vol. 369(9569):1302–13.

- Patel, V., and Kleinman, A. 2003. "Poverty and Common Mental Disorders in Developing Countries." *Bulletin of the World Health Organization*, Vol. 81(8):609–15.
- Razykov I, Ziegelstein RC, Whooley MA, & Thombs BD. 2012. "The PHQ-9 versus the PHQ-8- is item 9 useful for assessing suicide risk in coronary artery disease patients? Data from the Heart and Soul Study." *Journal of Psychosomatic Research*, Vol. 3: 163-168
- Rosselló, J., Bernal, G., & Rivera-Medina, C. 2008. Individual and group CBT and IPT for Puerto Rican adolescents with depressive symptoms. *Cultural Diversity and Ethnic Minority Psychology*, Vol.14(3), 234-245.
- Rosenberg, M. 1965. *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- Stein, M.B., and Kean, Y.M. 2000. "Disability and Quality of Life in Social Phobia: Epidemiologic Findings." *American Journal of Psychiatry*, Vol.157: 1606–13.
- Stillman, S., McKenzie, D., & Gibson, J. 2009. "Migration and Mental Health: Evidence from a Natural Experiment." *Journal of Health Economics*, Vol. 28(3):677–87.
- Tait, R.J., French, D.J., & Hulse, G.K. 2003. "Validity and Psychometric Properties of the General Health Questionnaire-12 in Young Australian Adolescents." *Australian and New Zealand Journal of Psychiatry*, Vol. 37(3):374–81.
- Vos, T. et al. 2012. "Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010." *The Lancet*, Vol. 380(9859), 2163 – 2196.
- Vos, T. et al. 2016. "Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015." *The Lancet*, Vol. 388(10053), 1545 – 1602
- World Bank. 2018. World Development Indicators. <https://data.worldbank.org/country/uganda>, accessed on September 17, 2019.
- World Health Organization (WHO). 2014. *Health for the world's adolescents: a second chance in the second decade of life*. Geneva: WHO.
- World Health Organization (WHO). 2011. Mental Health Atlas 2011. https://www.who.int/mental_health/evidence/atlas/profiles/uga_mh_profile.pdf?ua=1&ua=1, accessed on September 17, 2019.
- Zhou, S.G., Hou, Y.F., Liu, D., Zhang, X.Y. 2017. "Effect of Cognitive Behavioral Therapy Versus Interpersonal Psychotherapy in Patients with Major Depressive Disorder: A Meta-analysis of Randomized Controlled Trials." *Chinese Medical Journal*, Vol. 130: 2844-51

6. Appendices

6.1. Appendix A: Mentor Criteria

Mentors are selected from the same community within which the ELA club is located. In a new intervention community, the program assistant (PA) from the BRAC Uganda branch overseeing the ELA club conducts an initial community discussion (with adults and adolescent girls) to explain ELA model. During the discussion, she asks for their ideas where a club space can be rented and who could be potential mentors. There are a few guiding criteria for the community to give their nominations. The operational manual for the program lists eight main criteria. PA does the initial assessment and proposes the mentors, and the Area Coordinator makes the decision. A few of the criteria are subjective assessments of the assessors. Criteria #6 is the most binding one and the rest are guidelines (i.e. not as strictly followed). (The bold part is how it is shown in the manual)

1. She should be between 19 and 22 years old.
2. She must be a permanent resident of the village/community
3. She should be willing to work for four hours a day (club opening hours)
4. She should have the leadership quality, commitment and interest to work for adolescents.
5. She should be socially accepted in the community.
6. She **must not** be a student
7. She should be Ready to attend training and monthly refresher courses
8. She should have the capacity to run the club and other related activities effectively and efficiently

In addition, preference is given to those who have leadership experience either at school or in the community. It is also preferred that the mentor has a qualification of Secondary-2 and not a member of any microfinance group.

6.2. Appendix B: Group Based Interpersonal Psychotherapy (IPT-G)

6.2.1. IPT-G Background

The American Psychiatric Association, the American Psychological Association (APA, 2010), and the National Institute for Health and Clinical Excellence in the UK recognize IPT as an efficacious psychotherapy (British Psychological Society, 2010). There are now over 250 empirical studies supporting the efficacy and effectiveness of IPT.

In 2015 the WHO's Guideline Development Committee recommended that psychotherapy with or without pharmacotherapy may be used in the first line treatment of moderate to severe depression as first line treatment. This recommendation was made for high-income countries and especially for low to middle income countries. In 2016 the WHO released the *"mbGAP intervention guide for mental, neurological and substance use disorders in non-specialized health settings"*. On page 6 the guidelines the WHO makes the following recommendation: "As first-line therapy, health-care providers may select psychological treatments (such as behavioural activation, cognitive-behavioural therapy [CBT], and interpersonal psychotherapy [IPT]) or antidepressant medication (such as selective serotonin reuptake inhibitors [SSRIs] and tricyclic antidepressants [TCAs]). They should keep in the mind the possible adverse effects associated with antidepressant medication, the ability to deliver either intervention (in terms of expertise, and/or treatment availability), and individual preferences ... Different [psychological] treatment formats for consideration include individual and/or group face-to-face psychological treatments delivered by professionals and supervised lay therapists" (WHO, 2016).

The WHO also recommends that IPT or CBT be used as first line treatment in pregnant and breastfeeding mothers with moderate to severe depression, and also in adults with mild depression. In these groups anti-depressants should be avoided (WHO, 2016).

6.2.2. IPT-G: An Integrative Therapy

Klerman, Weissman, and colleagues (1984) devised IPT to optimize and leverage an array of change factors, including "common factors" of psychotherapy (Frank and Frank, 1991). IPT explicitly endeavors to instill hope and enhance expectation for change (Frank, 1971). Through use of the medical model, IPT seeks to create a new narrative for the client, demystifying and externalizing the current problem as something the client has rather than a defining aspect of who s/he is. Through use of the sick role, IPT seeks to decrease demoralization and guilt due to past social failures and the burden of current expectations, increase motivation for change (as it is the role of the client to now

get well), and emphatically validate the client's current distress. IPT explicitly values and builds on the supportive role of the therapeutic relationship.

Interpersonal therapy is a short-term, focused treatment for depression. Studies have shown that IPT, which addresses interpersonal issues, may be at least as effective as short-term treatment with antidepressants for mild to moderate forms of clinical depression. Events surrounding interpersonal relationships do not cause depression. But depression occurs within an interpersonal context and affects relationships and the roles of people within those relationships. By addressing interpersonal issues, interpersonal therapy for depression puts emphasis on the way symptoms are related to a person's relationships, including family and peers.

The immediate goals of treatment are rapid symptom reduction and improved social adjustment. The targets of IPT are symptom resolution, improved interpersonal functioning, and increased social support. Typical courses of IPT range from 6-20 sessions with provision for maintenance treatment as necessary. The defining elements of IPT can best be understood by describing the framework for its delivery. This framework can be divided into the theories supporting IPT; the targets of IPT; the tactics of IPT (i.e., the concepts applied in the treatment); and the techniques of IPT (i.e., what the therapist says or does in the treatment). Though individual elements in each of these categories may be shared with other psychotherapeutic approaches, their unique combination defines IPT.

IPT identifies one or more of four main problems, or triggers, that are associated with the onset of depression: Grief due to loss, Disputes or unsolvable disagreements with someone important, Life transitions (both good and bad), and Loneliness and social isolation. IPT helps people to explore, identify and express both positive and negative feelings about their problems. The therapy is specifically designed to focus on problems that are affecting the client's present life circumstances (i.e. the 'Here and Now') rather than focusing on the impact of past challenges or life experiences (Weissman, 2000). Throughout a 12-15-week IPT group therapy cycle, group members are encouraged to examine and seek further understanding about the link between their mood, events in their interpersonal lives to their moods and to turn to fellow group members for guidance. Through this introspection, combined with the practice of new behavior patterns within the group and in homework assignments, people find new ways to deal with the recurrent or persistent problems that are triggering their depression.

6.2.3. Understanding IPT Triggers

Grief and Loss: Related to depressive symptoms that started around the time of death of a loved one or in the following months. In a normal expression of grief, depressive symptoms go away in a few months. With unresolved grief symptoms are prolonged in an unhealthy way and can become a cause for impairment in someone's life. Using specific perspective taking exercises, IPT aims to help a person who is experiencing prolonged grief to fully mourn the loss of their loved one. The goal is to end the unhealthy perpetuation of grief and to make room in a person's life for other activities and people who will help them to enjoy life again.

Disputes or Unsolvable Disagreements: Depressive symptoms are connected to an ongoing disagreement with someone important in the person's life and may have to do with different expectations between two people close to each other about how to handle situations or about what each person wants from the other. Some of the hopelessness is also related to a core belief that the disagreement cannot be solved. IPT asks the person to look at the disagreement using the following three stages: **Renegotiation** – When the person is still trying to resolve the disagreement. In IPT, the facilitator and the group help the person find different ways of talking to the other person to manage the problem. A person in this stage still wants to work things out but needs help with how to do this. **Impasse** – When the person feels like everything has already been tried and nothing will work. The person feels stuck. Talking has stopped which has resulted in a lot of residual anger. In IPT, the facilitator and group members try to get the person to “try one more time,” and to find new ways of handling the problem (including potentially letting it go). **Dissolution** – When the person is at the end of the relationship and it is too late for anything to get better. In IPT, the facilitator and the group help the person prepare emotionally and physically to end the relationship. When facing disputes or unsolvable disagreements, IPT aims to help someone decide what they want in the relationship, to figure out their stage of disagreement, and to develop, practice and implement new skills for talking to the other person in the disagreement about what they want and need in order to reach resolution.

Life Changes and Transitions: If a person enters depression around the genesis of a life change then often they are having difficulty managing the new situation and they do not feel prepared to handle the new parts of their life. Adjusting to life changes means understanding, accepting and fully integrating the losses and gains that the new situation presents. In IPT, the therapeutic goal is to help a person recognize their feelings about the change (such as sadness, anger, confusion or powerlessness), then to help them see what might be positive about the change and finally to help them identify and learn the skills necessary so that they can fully respond to and manage the difficult aspects of the change.

Loneliness and Social Isolation: Depression related to loneliness or social isolation can occur in people who have a history of problems making friends and/or keeping them or who express an innate feeling of separation from others. Unlike the 'here and now' aspect of the other triggers, people who experience depression related to loneliness or isolation report that the feelings are not connected to anything that has happened recently. These deep-seated emotions have been felt for a long time. IPT and the group dynamic are used to gain understanding for the basis of this belief system. The goal is to identify and understand the root belief and/or the root behavior that is resulting in a pattern of isolation. Then, with a new way of understanding their thoughts or actions, the person can practice, within the group and through homework exercises, making friends by learning new approaches to start and keep friendships (Weissman, 2000).

Group therapy sessions build bonds between women and encourage them to actively engage in the healing process and to support each other in the exploration of their depression triggers. With new healthier patterns and skills, women can learn to manage their current depression and ensure future depressive episodes can be quickly identified and resolved before the onset of any long-term consequences.

6.3. Appendix C: StrongMinds Uganda (SMU) Delivery of IPT-G

IPT-G is delivered in four phases namely pre-group, initial phase, middle phase and termination phase. The pre-group phase takes place before the actual group therapy sessions start. Facilitators hold one on one session with potential clients and re-administer the PHQ-9²⁹ to confirm diagnosis and discuss possible triggers.

The initial phase takes place over the first five weeks of therapy. It involves working with the clients to learn the relationship between events and their mood, and the relationship between events – mood – trigger/problem area. During this phase clients may identify one or more problem area that they will want to work on during the course of therapy. The middle phase takes place between week 6 and 12. During this phase the clients test different ways of breaking the link between event-mood and trigger by focusing on solving or resolving the problem areas. Those that identified grief work on developing skills for healthy grieving, those that identified unresolved disputes work on skills to solve the dispute, resolve why they are stuck at an impasse or end the relationship; those who identified life change develop skills to adjust to the new life and those who identified loneliness develop skills to create or start new human connections. The termination phase take place between week 13 and 14 and during this phase the facilitator and clients celebrate reduction of symptoms, skills developed and tested to solve problem areas and new friendship formed.

89 mentors were trained in three batches over the course of three weeks. Each training lasted six days. The objectives of the training were to:

1. Build general counselling skills
2. Improve mentor's understanding of depression including adolescent depression.
3. Introduce the mentors to the IPTG technique
4. Develop skills on how to conduct a pre-group session
5. Develop skills on how to conduct the first 10 session of IPTG.

The training was conducted by four facilitators and covered basic counselling skills, overview of depression, introduction to IPT-G, the pre-group phase and the initial and middle phases of IPTG. These are each summarized in more detail below:

- i. *Counselling skills* – The training was delivered through lectures and role play sessions. The topic delivered included basic counselling skills (active listening, probing and questioning, sensitivity

²⁹ Note that the PHQ-9 will be used during therapy, while the PHQ-8 will be used by the research team

and the role of the counsellor), ethics of counseling (confidentiality, empathy, non-judgmental etc.), why I need supervision.

- ii. *Overview of depression* – defining or describing depression, symptoms of depression and how to diagnose depression using PHQ – 9.
- iii. *Introduction to IPT-G* – The training introduced IPT-G and its unique approach to counselling. This included introducing the four problem areas and the link between daily events, mood and problem areas.
- iv. *How to conduct a pre-group session* – The training focused on delivering the key tasks of a pre-group session including role play and practice sessions. The Pre- group is a one on one session between the facilitator and the potential client. At the end of Pre-group, the client should understand that they have an illness called depression, the client will explore which of the four problem areas may have triggered or exacerbated the depression, be encouraged about the possibility of recovery as a result of attending group therapy sessions and set a goal for recovery. The key tasks of pre-group include screening for depression, screening for suicide risk her a client responds to question 9 on PHQ-9. Initiating a safety plan for clients who score above 5 on the suicide questionnaire, referral for further treatment for clients who score above 11 on the suicide assessment, probing for problem areas/triggers, and goal setting. Please note that the mentors were trained to assess for suicide risk but it was the supervisors who conducted the safety planning and the referral to specialized mental health services.
- v. *How to conduct the initial and middle phases of IPTG* – the training focused on delivering the key tasks of the initial and middle phase. The trainers provided in-depth training how to apply the four problem areas of IPTG in group therapy. The continued to build on the use of basic counselling skills and ethics of counselling.

SMU plans to have one or two refresher trainings in the course of the 14-week therapy coinciding with the end of the initial and middle phases. In addition to this SMU Supervisors will provide weekly training during debrief sessions help at the branches. The weekly training will ensure that (i) Mentors are ready to deliver the therapy for the new week and (ii) Take corrective action based on areas of weakness identified during the observation of sessions. Future support for the duration of therapy includes:

- i. *Scheduled and impromptu supervision visits.* SMU Mental Health Supervisors (MHS) will conduct both scheduled and impromptu supervision visits to observe the mentor at work. They will assess the mentor using criteria laid out in the SMU quality assurance tool. They will provide

immediate feedback to the mentor at the end of the session and extra training may take place on site.

- ii. *Debriefing session.* All BRAC branches are attached to a SMU MHS. They have divided their time in a manner that will allow for them to hold weekly debriefing session with the mentors attached to each branch. The debrief session is divided into three parts: (i) To review the past week, review mentor files and collect session specific data (group attendance and burden rating); (ii) MHS provides feedback from the week’s observation, checking whether any of the other mentors are struggling with issues she observed in the field. Corrective action will be taken immediately through training, role play session etc.; and (iii) MHS conducts training on the session for the upcoming sessions and supports the mentors to make session plans.
- iii. *Data entry.* – SMU will enter therapy related data into its M&E system and provide the relevant analysis at the end of the cycle.

6.4. Appendix D: Survey Modules and Baseline Balance

Appendix Table D1: Survey Modules

Module	Brief description
Introduction and screening	General information, informed consent, screening (PHQ-8)
Social network integration	Information on the respondent’s friendships and feelings of social integration
ELA club	Information on current and/or former attendance of ELA clubs
Family background	Information on the respondent’s household head, biological mother and father
Poverty Probability Index (PPI)	Information on household’s characteristics and asset ownership which are scored to compute the likelihood that the household is living below the poverty line
Assets	Ownership of additional productive assets, including land owned and livestock
Education	Information of the respondent’s current school attendance and qualifications, as well as vocational training and apprenticeship

Psychosocial wellbeing (GHQ-12)	Information on self-reported psychological disorders and strains as evaluated with the 12-item General Health Questionnaire
Time allocation	Hourly recall of main activities the respondent carried out on the day before
Self esteem	Information on the respondent's confidence in her worth or abilities scored through the Rosenberg Self Esteem Scale
Income-generating activities	Information on paid activities, recent income, and job opportunities for young people
Financial inclusion and economic empowerment	Information on the respondent's access to money, credit and savings.
Resilience	Information on the respondent's availability of resources that may bolster her resilience evaluated through a revised version of the Child and Youth Resilience Measure
Marital status and history	Information on the respondent's current and past relationships, and future marriage expectations.
Sexual history	Information on the respondent's sexual activity, previous partners and contraceptive use
Fertility	Information on the respondent's birth history and future pregnancy expectations
Violence	Information on the respondent's exposure to emotional, physical, financial and sexual violence
Competencies	Assessment of the respondent's skills in Math, English/Luganda literacy and comprehension, and ability to follow instructions

Appendix Table D2: Baseline Means and Balance

	Control Group Mean (s.d.)	Baseline Difference (s.e.)		p-value (IPTG- IPTG+)
		IPTG	IPTG+	
	(1)	(2)	(3)	(4)
<i>Primary Outcomes</i>				
Patient Health Questionnaire-8 Score (0-24)^	13.200 (3.000)	-0.003 (0.134)	0.016 (0.158)	0.900
=1 if score on General Health Questionnaire-12 >=3	0.934 (0.248)	-0.006 (0.015)	0.009 (0.012)	0.257
Rosenberg Self Esteem Scale Score (0- 36)	16.280 (3.781)	-0.163 (0.258)	-0.340 (0.256)	0.499
Child and Youth Resilience Measure Score (0-34)	25.188 (5.428)	-0.651 (0.515)	0.146 (0.522)	0.079
<i>Secondary Outcomes and Covariates</i>				
=1 if Enrolled in School	0.416 (0.493)	0.012 (0.032)	-0.029 (0.026)	0.210
Number of Correct Answers on Competencies (0-4)	1.309 (1.465)	0.008 (0.095)	0.034 (0.089)	0.782
=1 if Never Married	0.871 (0.336)	0.012 (0.022)	0.026 (0.019)	0.469
=1 if Never Pregnant	0.816 (0.388)	0.001 (0.027)	0.027 (0.024)	0.298
=1 if Risky Sex	0.230 (0.421)	-0.017 (0.028)	-0.046 (0.025)	0.292
Age of Adolescent	16.760 (2.026)	-0.039 (0.144)	-0.003 (0.144)	0.794
Score on Poverty Probability Index (0- 100)	56.756 (10.594)	1.264 (0.786)	0.462 (0.769)	0.343
Chi-squared joint test of orthogonality (p-value)		0.240	0.342	0.271

Notes: Minimal depression as scored by the PHQ-8 is zero for all participants (by design) so we show the PHQ-8 score for completeness. For PHQ-8 and GHQ-12, higher scores indicate worse outcomes, while for self-esteem, resilience, competencies and poverty probability index higher scores indicate better outcomes. For Column (1) shows the mean and standard deviation in the control group. Column (2) shows the baseline difference between IPTG and control, with column (3) showing the difference between IPTG+Cash and control. Column (4) presents the p-value on the difference between the two treatment groups. Standard errors are clustered at the community level. Mean differences statistically different than zero at 99% (***) , 95% (**), and 90% (*) confidence.

7. Administrative Information

7.1. Funding

This work was supported by the Wellspring Advisors, LLC [grant number 5987] and the Center for Effective Global Action who received funding from an anonymous donor.

7.2. Institutional Review Board

The study design was approved by the George Washington University Committee on Human Research, Institutional Review Board (180580), Makerere University College of Health Sciences School of Public Health Higher Degrees, Research and Ethics Committee (552), and the Uganda National Council for Science and Technology (HS318ES). The proposals submitted to these entities provide detailed documentation of the ethical considerations of this study including a detailed protocol if an adolescent expressed suicidal ideation. These documents are available upon request.

7.3. Declaration of interest

Baird, Özler and Dell’Aira have no conflicts of interest to declare. Us Salam was employed by the BRAC Independent Evaluation and Research Cell Unit (IERC) Uganda during baseline data collection. His performance was not evaluated on the results of the impact evaluation, and BRAC IERC has developed a reputation of publishing positive and negative findings. Us Salam no longer works at BRAC IERC.

7.4. Acknowledgements

We gratefully acknowledge support in developing the original proposal from Munshi Sulaiman (BRAC IERC Uganda), Temina Madon (CEGA), and Adam Schwartz (BRAC USA). In addition, we acknowledge efforts by StrongMinds (Kari Frame, Dena Beatrice and Christina Ntulo) and BRAC Uganda (Naume Abunyo) to adapt programming for this project and to ensure that implementation maintained the fidelity of the research design.

Author Contributions

Sarah Baird: Conceptualization, Methodology, Investigation, Formal Analysis, and Writing-Original Draft preparation. **Berk Özler:** Conceptualization, Methodology, Investigation, Formal Analysis, and Writing-Original Draft preparation. **Chiara Dell’Aira:** Data Curation, Investigation, Project Administration, Writing-Review and Editing. **Danish Us’Salam:** Data Curation, Investigation, Project Administration, Writing-Review and Editing.