Pre-Analysis Plan for:

Mental Health, Productivity, and Child Investment in Bangalore

Manuela Angelucci*
Daniel Bennett†

May 4, 2017

1 Introduction

This study evaluates the impact of depression treatment and job placement assistance on mental health and labor force participation in Karnataka, India. Depression is pervasive in developing countries and may have widespread but unexamined economic consequences by affecting labor supply, child investment, and risk and time preferences. Correlational studies from developed countries link mental disorders to poor job performance (Kessler and Frank 1997, Berndt et al. 1998, Fletcher 2013). This study will provide the first causal evidence in any setting of the impact of depression on labor market outcomes. We will also provide an important contribution to the literature on the impact of parental depression on child outcomes (Cummings and Davies 1994, Beck 1998, Letourneau et al. 2013). These mechanisms are important for economic development because they may foster poverty traps (Case and Deaton 2005, Banerjee and Duflo 2012).

We completed this pre-analysis plan before fielding the midline survey. Below we detail the study design, specify our primary outcomes, and describe our econometric strategy.

2 Study Design

2.1 Interventions

Grameena Abhyudaya Seva Samasthe (GASS) is a social service organization based in Doddaballapur, Karnataka. The organization works to rehabilitate people with physical and

*Department of Economics, University of Michigan, mangeluc@umich.edu.
†Center for Economic and Social Research, University of Southern California, bennettd@usc.edu

mental disabilities by providing health care and "livelihoods assistance." We collaborated with GASS to offer depression treatment and job placement assistance to adult community members.

- *Depression Treatment Intervention*: GASS organized monthly mental health camps in each taluk. Participants received a diagnostic visit and seven follow-up visits over eight months. Psychiatrists diagnosed and treated patients based on their professional assessments. As appropriate, they prescribed established antidepressants, which we acquired generically through reputable Indian manufacturers. Psychiatrists monitored progress and side effects throughout the intervention and adjusted drugs and dosages as needed. GASS volunteers visited participants at home every month throughout the intervention to monitor mental health, drug compliance, and side effects.

- *Job Placement Intervention*: GASS offered job placement support and attempted to place participants in formal sector jobs. Many of these jobs were in local factories making garments, plastics, medicine, wristwatches, paint, and cosmetics. These jobs paid a minimum of Rs. 6000 per month and complied with the labor laws of Karnataka. For participants who were not interested in these jobs, GASS worked to identify and facilitate other income-generating activities. GASS organized monthly group meetings to discuss work-related challenges and encourage retention.

- *Depression Monitoring Intervention*: GASS monitored the mental health of participants who did not receive depression treatment. Staff members visited participants at home on a monthly basis and administered the PHQ-9 to monitor depression symptoms.

2.2 Recruitment and Participation

We conducted the study in the taluks of Doddaballapur, Korategere, and Gouribidanur, which are 14-40 kilometers north of Bangalore, Karnataka. These taluks include 760 villages and wards with a 2011 population of 194,000. GASS identified 616 villages and wards where it was feasible to offer the interventions.

Surveyors conducted a door-to-door screening campaign within the sample villages. They followed a pre-determined door-skip pattern based on village size (in the 2011 census) with a target of enrolling two eligible participants per village. They screened people for depression using the PHQ-9 instrument (Kroenke et al. 2001). They also sought to identify people who would benefit from the job placement intervention. We required participants to meet the following eligibility criteria: (a) has a PHQ-9 score of 7-20, (b) aged 18-50, (c) not pregnant, (d) not currently earning more than Rs. 6000 per month, (e) no child care duties, (f) has family permission to work. Surveyors screened 6446 people to identify 1000 study participants.

We monitored the severity of depression for all study participants throughout the study. Psychiatrists monitored the mental health of participants in the Depression Treatment intervention, while GASS staff conducted monthly home visits to participants in the Depression Monitoring intervention. We immediately removed from the study and referred for treatment anyone who scored ≥ 21 on the PHQ-9 during these contacts.

2.3 Experimental Methodology

We randomized villages into four intervention arms before participant recruitment. We randomized by village to minimize communication among participants receiving different interventions.

- T1: depression treatment and job placement assistance (102 villages)
- T2: depression treatment only (94 villages)
- T3: job placement assistance and mental health monitoring (104 villages)
- T4: mental health monitoring only (202 villages)

We carried out the randomization in our offices using Stata. We allocated roughly 20 percent of villages to each of T1, T2, and T3, and 40 percent of villages to T4. This approach enhanced power given a constraint on the number of job placements GASS could feasibly make within the study interval.

We stratified the randomization by taluk and an index of village socioeconomic status derived from the 2011 census. We constructed this index by computing the first principal component of the following village characteristics: percent of households with a high-quality roof, a high-quality floor, electricity, a latrine, a radio, a television, internet access, a computer, a landline phone, a mobile phone, a bicycle, a motorized scooter, and a car, as well as the mean number of rooms per dwelling. Then we divided the socioeconomic status index into terciles, which led to a total of $3 \times 3 = 9$ strata for randomization.

2.4 Surveys

We completed the screening and baseline survey from December 2016 - March 2017. Surveyors immediately conducted baseline surveys with participants who were eligible and provided informed consent to participate. We will conduct a midline survey in May - June of 2017 and an endline survey in October - November of 2017. We chose to conduct two follow-up surveys to enhance statistical power in regressions that pool the observations from these rounds.

The survey includes five incentivized tasks: (a) a lentil sorting task to elicit time preferences, (b) a lottery choice task to elicit risk preferences, (c) an 8-item battery of Ravens Progressive Matrices, (d) a forward-digit recall task, and (e) a backward digit recall task. After completing these tasks, participants roll a five-sided die to select which task will be incentivized. Participants can earn between Rs. 0 - 200 through these tasks, depending on their responses.

3 Plan for Data Analysis

3.1 Estimation Strategy

We will estimate the causal effect of depression treatment (DT), job placement assistance (JP), and the interaction between these interventions on mental health and labor market

outcomes. Equation (1) is an ANCOVA specification and Equation (2) is a differences-in-differences specification. The efficient specification depends on the degree of serial correlation in the dependent variable. We will follow McKenzie (2012) and use ANCOVA for outcomes with low serial autocorrelation and differences-in-differences for outcomes with high autocorrelation.

$$Y_{ijt} = \alpha_0 + \alpha_1 DT_j + \alpha_2 JP_j + \alpha_3 [DT_j \cdot JP_j] + \alpha_4 Y_i^b + \alpha_5 X_{ij}^b + \delta_t + \varepsilon_{ijt} \quad \text{for } t > 0 \tag{1}$$

$$\begin{aligned} Y_{ijt} = &\beta_0 + \beta_1 [DT_j \cdot P_t] + \beta_2 [JP_j \cdot P_t] + \beta_3 [DT_j \cdot JP_j \cdot P_t] + \beta_4 DT_j + \beta_5 JP_j \\ &+ \beta_6 [DT_j \cdot JP_j] + \beta_7 X_{ij}^b + \delta_t + \varepsilon_{ijt} \quad \text{for } t \geq 0 \end{aligned} \tag{2}$$

i indexes the individual, j indexes the village, and $t \in \{0, 1, 2\}$ indexes the survey round in these equations.[1] Y is the outcome variable, Y^b is the baseline dependent variable, X^b is a vector of baseline covariates, P is a "post" indicator, and δ is a vector of round dummies.

We will also estimate a specification that omits the interaction between DT and JP. This specification maximizes the power to detect direct effects of the interventions but does not identify the complementarity between these interventions. We will also estimate specifications that distinguish between impacts in the midline and endline survey rounds.

3.2 Multiple Inference Correction

This study examines impacts on many outcome variables. Depression plausibly influences many different aspects of economic behavior and decision making, and our study will provide among the first causal evidence of these impacts in practice. We divide outcome variables into four categories: (1) primary outcomes, (2) mechanism outcomes, (3) exploratory outcomes, and (4) household spillover outcomes. These categories include subcategories that capture multiple ways of measuring the same concept. We may analyze results across multiple papers.

We will deal with multiple inference in two ways. Where possible, we will create indices that collapse outcome variables into a single dependent variable. Alternatively, we will report p-values that are adjusted for the false discovery rate (FDR) using the Romano and Wolf (2005) stepdown procedure. We will adjust p-values across outcomes within subcategories.

3.3 Heterogeneity

We will assess treatment effect heterogeneity according to several baseline characteristics. We will estimate treatment effects by subgroup for dichotomous variables. For continuous variables, we will assess the statistical significance of heterogeneous treatment effects non-parametrically. For exposition, we may divide continuous variables into categories that capture relevant variation.

- Age
- Gender

[1] Some outcomes such as consumption pertain to the household as a unit. Other outcomes such as child health and education pertain to other household members.

- Socioeconomic status: we will compute the first principal component of the following baseline variables: completed schooling, literacy, monetary savings, dwelling characteristics (number of rooms, quality of roof, quality of floor) ownership of land and durable goods (chair, bed, table, electric fan, television, refrigerator, bicycle, motorcycle, car).

- Mental health: we will use the PHQ-9 depression severity score (Kroenke et al. 2001) and the GAD-7 anxiety severity score (Spitzer et al. 2006).

- Physical health: we will compute the first principal component of several activities of daily living: the ability to (a) carry out vigorous activities, (b) carry out moderate activities, (c) bathe and dress without help, (d) carry a 10-kilogram object for 500 meters. The index will also include (e) the number of kilometers the respondent can walk without getting tired, (f) degree of physical pain, and (g) the degree to which pain has interfered with daily activities.

- Early-life shocks: the baseline survey measures the exposure to six shocks before the age of 16 (death of a parent, death of another family member or close friend, a serious illness for the respondent, a serious illness for a family member, a situation of emotional or physical violence, serious financial problems). We will aggregate these outcomes by computing the number of shocks the individual has experienced. We will weight shocks by severity, following Holmes and Rahe (1967).

- Cognition: the baseline survey includes eight Ravens Progressive Matrix responses, as well as forward digit recall and backward digit recall assessments. We will compute the first principal component of the number of correct Ravens responses, and the two digit recall scores.

3.4 Experimental Integrity

We will assess experimental integrity by comparing the baseline characteristics of the intervention arms at baseline. If we have carried out the randomization correctly, fewer than 5 percent of these comparisons should be significant at the 5-percent threshold. We will measure significance by regressing each outcome on indicators for T1, T2, T3, with T4 as the excluded group. We will jointly test the significance of T1, T2, and T3. We will cluster standard errors by village for this exercise. We will conduct this exercise using both the baseline and endline samples (i.e., the subsample of non-attriters).

We will assess balance for all of the outcome variables listed below. We will also assess balance for the following baseline covariates: (a) age, (b) gender, (c) literacy, (d) household asset index, (e) housing quality index, (f) exposure to early-life shocks, (g) household size, (h) presence of open defecation, (i) presence of garbage, (j) cleanliness of kitchen, (k) exposure to early-life shocks, (l) subjective well-being.

4 Outcomes

4.1 Primary Outcomes

- Depression severity: the PHQ-9 depression severity scale incorporates nine items and ranges from 0 to 27.
- Time use: we will measure the following activities over 1-day and 7-day intervals. The 1-day interval measurements are based on a 24-hour activity diary while 7-day interval measurements are derived from lists of paid and unpaid jobs.
  - Time devoted to paid work. In the activity diary, we isolate "working at main job", "working at secondary job", and "farming, gardening, or animal husbandry." In the job list, we add the hours worked across up to three jobs.
  - Time devoted to domestic work. In the activity diary, we isolate "cooking and preparing meals", "cleaning kitchen", "laundry", and "fetching water."
  - Time devoted to child care. In the activity diary, we isolate "caring for children". We also directly elicit the number of hours per week spent caring for children younger than 13 in the past 7 days.
- Earnings by the respondent over 1-week and 1-month intervals. We compute earnings over the past 7 days for up to three jobs from the job list. We also elicit the respondent's earnings in the past month.

4.2 Mechanism Outcomes

Impacts on these outcomes may help explain findings related to the primary outcomes above.

- Time preferences. We include an incentivized lentil sorting task that elicits the time preference for leisure with the "convex time budget" methodology (Andreoni and Sprenger 2012). Participants are asked to sort a mixture of brown and yellow lentils in exchange for Rs. 200. They may decide whether to sort lentils in an earlier or later period, and the total quantity sorted depends on the allocation of work between periods. We use three combinations of time periods and five alternative interest rates, for a total of 15 choices. All participants respond but only a fraction actually sort lentils.
  - We will use the responses to compute the curvature of the utility function, the weekly discount factor, and the degree of present bias (Carvalho et al. 2016).
  - We will estimate impacts on the percent of lentils sorted early using choice-level data. We will condition on the interest rate, the time interval, and the proximity of the choice to the present. We will also condition on the interaction between these variables and DT, JP, and $DT \cdot JP$ indicators.
- Risk preferences. We include three distinct risk preference elicitations. We will compute the first principal component of these measurements to create a risk preference index. We will estimate impacts on both the mean and the variance of the index.

  - General subjective risk attitude.
  - Modified DOSPERT Scale (Blais and Weber 2006): we assess attitudes about five risky behaviors (leaving young children alone for 30 minutes, riding a motorbike without a helmet, lending Rs. 1000 to a neighbor in a financial emergency, eating a food that might be spoiled when there is nothing else to eat, investing 10 percent of annual income in a new business venture, delaying a trip to the doctor for a child with diarrhea). In each case, we elicit how risky the respondent perceives the behavior and how likely he or she is to engage in the behavior. We will aggregate responses by weighting the willingness to take each action by the perceived riskiness of each action.
  - Incentivized lottery choice: respondents choose across six lotteries that have different means and variances (Eckel and Grossman 2008). Payoffs range from Rs. 5 to 175.

- Cognition: we will compute the first principal component of the following three cognition metrics. We will also consider each metric separately. In each case, participants who randomly select the task will receive Rs. 15 per correct response.
  - Ravens Progressive Matrices: the survey includes an eight-item battery. The specific items will change across survey rounds. We will compute the number of correct responses.
  - Forward digit recall: participants will attempt to recall up to eight digits in forward order. The task ends once the participant makes a mistake. We will record the number of digits they recall correctly.
  - Backward digit recall: participants will attempt to recall up to eight digits in backward order. The task ends once the participant makes a mistake. We will record the number of digits they recall correctly.

4.3 Exploratory Outcomes

There is little evidence about the impacts of depression and depression treatment on economic outcomes. We measure many possible channels, which will enable us to explore possible linkages.

- Anxiety: the GAD-7 anxiety severity scale includes seven items and ranges from 0 to 21 (Spitzer et al. 2006). We will also consider the impact on depression and anxiety comorbidity.
- Household food consumption: we measure the amount of food purchases, grown, or received for free in the past seven days. Surveyors ask specifically about 23 food categories. The survey includes the quantity and expenditure, as well as the share that the respondent personally consumed.
- Individual food consumption: the portion size, cost, and composition (across 23 food categories) of the previous night's meal.

- Bargaining power: we ask which household member (a) decides who may work outside the home and (b) whether to save for the future. We will create a variable that indicates how many of these decisions the respondent participates in.

- Subjective wellbeing: respondents are asked how strongly they agree or disagree with the following statements: (a) in most ways my life is close to ideal, (b) the conditions of my life are excellent, (c) I am satisfied with my life, (d) so far I have gotten the important things I want in life, and (e) if I could live my life over, I would change almost nothing. We will create an index by computing the first principal component of these responses.

- Physical health: we will compute the first principal component of several activities of daily living: the ability to (a) carry out vigorous activities, (b) carry out moderate activities, (c) bathe and dress without help, (d) carry a 10-kilogram object for 500 meters. The index will also include (e) the number of kilometers the respondent can walk without getting tired, (f) degree of physical pain, and (g) the degree to which pain has interfered with daily activities.

- Additional time use outcomes: time allocated to sleep, leisure, and job search within the past 24 hours.

- Health technology adoption: during the midline survey, we will market a 100ml bottle of hand sanitizer to participants.

  - To begin, surveyors will provide a brief hygiene lesson.
  - We will elicit willingness to pay for a 100ml bottle of hand sanitizer using the Becker, DeGroot and Marschak (1964) algorithm. In this procedure, respondents draw prices from a bag and must purchase the good if their bid exceeds the price that they draw. We will set all of the prices in the bag to be zero so that everyone receives the 100ml bottle for no cost. We will also provide participants with a 500ml bottle.
  - We will randomize participants into this intervention by village. 80 percent of respondents will receive the hygiene intervention.
  - We will measure hand sanitizer use at follow-up by inspecting the bottles provided at midline. We will also elicit self-reported use.
  - The midline and endline surveys include a brief assessment of hygiene knowledge, which will allow us to examine mechanisms related to this intervention.

4.4 Household spillover outcomes

The outcomes below allow us to assess possible impacts of the intervention on other household members.

- Consumption

  - Household food consumption: the survey measures the consumption and expenditure on 23 food categories within the past week.
  - Household non-food expenditures: the survey measures spending across 10 categories (communications, utilities, rent, tobacco, toiletries, entertainment, medication, doctor fees, other medical expenses, and religious activities). It also measures clothing expenditures for both adults and children. We will examine impacts on:
    * Total household expenditures
    * Medical expenditures
    * Expenditures on clothing for children

- Asset ownership: household ownership of durable goods (chair, bed, table, electric fan, television, refrigerator, bicycle, motorcycle, car). We will compute the first principal component of these items. We will also examine monetary savings.

- Sanitation: the survey includes surveyor observations of the degree of open defecation and garbage disposal around the household, as well as the cleanliness of the cooking area.

- Child health: respondent-reported diarrhea, fever, and cough within the past two weeks, breastfeeding, child weight and height for children under age 5 at baseline.

- Child education: school enrollment, attendance, and homework time for children aged 5-18 at baseline.

References

Andreoni, James and Charles Sprenger, "Estimating Time Preferences from Convex Budgets," *American Economic Review*, 2012, *102* (7), 3333–3356.

Banerjee, Abhijit and Esther Duflo, *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty* 2012.

Beck, C, "The Effects of Postpartum Depression on Child Development: a Meta-analysis," *Archives of Pediatric Nursing*, February 1998, *7* (1), 12–20.

Becker, Gordon, Morris DeGroot, and Jacob Marschak, "Measuring Utility by a Single-Response Sequential Method," *Behavioral Science*, July 1964, *9* (3), 226–232.

Berndt, E, SN Finkelstein, PE Greenberg, RH Howland, A Keith, AJ Rush, J Russell, and MB Keller, "Workplace Performance Effects from Chronic Depression and its Treatment," *Journal of Health Economics*, 1998, *17*, 511–535.

Blais, Ann-Renee and Elke Weber, "A Domain-Specific Risk-Taking (DOSPERT) scale for adult populations," *Judgment and Decision Making*, July 2006, *1* (1), 33–47.

Carvalho, Leandro, Stephan Meier, and Stephanie Wang, "Poverty and Economic Decision-Making: Evidence from Changes in Financial Resources at Payday," *American Economic Review*, 2016, *106* (2), 260–284.

Case, Anne and Angus Deaton, "Health and Wealth Among the Poor: India and South Africa Compared," *American Economic Review: Papers and Proceedings*, May 2005, *95* (2), 229–233.

Cummings, EM and PT Davies, "Maternal Depression and Child Development," *Journal of Child Psychology and Psychiatry*, 1994, *35* (1), 73–112.

Eckel, Catherine and Philip Grossman, "Forecasting risk attitudes: An experimental study using actual and forecast gamble choices," *Journal of Economic Behavior and Organization*, 2008, *68*, 1–17.

Fletcher, Jason, "Adolescent Depression and Adult Labor Market Outcomes," *Southern Economic Journal*, 2013, *80* (1), 26–49.

Holmes, Thomas and Richard Rahe, "The Social Readjustment Rating Scale," *Journal of Psychosomatic Research*, 1967, *11*, 213–218.

Kessler, RC and R Frank, "The Impact of Psychiatric Disorders on Work Loss Days," *Psychological Medicine*, 1997, *27*, 861–873.

Kroenke, Kurt, Robert Spitzer, and Janet Williams, "The PHQ-9: Validity of a Brief Depression Severity Measure," *Journal of General Internal Medicine*, September 2001, *16* (9), 606–613.

Letourneau, N, L Tramonte, and JD Willms, "Maternal Depression, Family Functioning, and Children's Longitudinal Development," *Journal of Pediatric Nursing*, 2013, *28*, 223–234.

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

Romano, Joseph and Michael Wolf, "Exact and Approximate Stepdown Methods for Multiple Hypothesis Testing," *Journal of the American Statistical Association*, March 2005, *100* (469), 94–108.

Spitzer, Robert, Kurt Kroenke, Janet Williams, and Bernd Lowe, "A Brief Measure for Assessing Generalized Anxiety Disorder," *Archives of Internal Medicine*, May 22 2006, *166*, 1092–1097.