Youth FORWARD
Data Analysis Protocol
Feb. 6, 2020

The Youth FORWARD Hybrid Type 2 Effectiveness-Implementation scale-up study proposes to implement and test an evidence-based intervention called the Youth Readiness Intervention (YRI) alongside the German Development Agency's (GIZ) Entrepreneurship Training program (ENTR) which is part of a larger youth Employment Promotion Programme (EPP). Youth who apply to the employment program are screened for eligibility and those youth who are deemed eligible are clustered and randomized into three study arms: ENTR-only, YRI+ENTR, Control. Below we outline our procedures for clustering youth as well as our detailed data analytic plan for addressing Study Aim 3, which regards clinical effectiveness of the YRI when delivered by an alternate delivery platform (ENTR).

Analytic procedures for cluster matching prior to randomization:

Using rules designed to prevent spillover/diffusion to villages outside of a cluster, two types of clusters were defined either standalone large towns, which could not be matched with comparable smaller villages, or, more commonly, groups of neighboring villages. Because the total number of clusters available for randomization was too small to assume comparability on covariates/confounders measured or unmeasured, we matched triads of "similar" clusters that could be randomized into ENTR, YRI+ENTR arms and a control arm. Data were available from participant applications to participate in the training opportunities and in the eligibility screening surveys, as well as known data about the municipalities (see list below).

To accommodate the requirements of the study we used a "threshold blocking" (Higgins, Sävje, & Sekhon, 2016), allowing the creation of groups of at least or exactly *n* matched items based on a distance measure. Following common practice, we define "similarity" on the basis of distance in a multidimensional space. In this study we used Euclidean distance as a measure of similarity, and we computed it using standardized variables to protect against scale effects.

Procedurally, we implemented the matching by first aggregating the individual-level data up to the cluster level. We produced proportions from binary variables as well as categorical variables with more than two levels (which we split into binary variables). We computed the means of continuous variables and kept indicator variables unchanged. We then used the following variables: age, sex, marital status, number of dependents, education, previous skill training, income generating activities, days and hours worked in the past month, WHO Disability Assessment Schedule (WHODAS) score, Difficulties in Emotion Regulation Scale (DERS) score, locality has access to highway, locality is a hub village, and locality population.

We stratified by variable standalone which flags standalone clusters in the data, and these large towns, were matched in triads separately using the same technique. Matching the significantly larger towns to smaller localities (villages) was deemed inappropriate. Finally, we used the algorithm of Higgins et al as implemented in R package scclust to group the clusters into groups of three, which were then randomized into sets of each of the ENTR, YRI+ENTR and the untreated control. This procedure could and did result in some unmatched clusters. These were separately randomized maintaining the best possible balance among the ENTR, YRI+ENTR and control conditions (e.g., for five unmatched clusters, 2-2-1) and are tagged with dummy variables both to assess the effect of the unmatched clusters on later analyses and to control for the imbalance through a fixed effects approach.

Because it is impossible even after matching pre-randomization and randomization to attain perfect balance across clusters on any one analytic variable, all analyses will include control for a cluster-level aggregate at baseline (pretest) of the outcome variable under investigation. Furthermore, prior to analysis, we will examine the clusters for imbalance for baseline levels of all outcomes as well as likely confounders. If there is imbalance on important confounders across clusters, aggregates of these confounders at baseline will be included at the cluster level during analysis.

Youth FORWARD scale-up study AIM 3:

AIM 3 (Clinical Effectiveness): To compare clinical effectiveness of YRI when delivered via the EPP platform to results of our previous randomized control trial (RCT) of YRI as measured by improved mental health and reduced functional impairments among high-risk youth. Emotion regulation will be examined as a major mechanism by which the YRI improves behavior of treated youth and their functioning in the ENTR.

> *Hypothesis 3(a)*: *Participants who are exposed to YRI will demonstrate greater reduction in mental health and functional problems than participants who do not receive the YRI intervention; emotion regulation (as measured using the Disturbances in Emotion Regulation Scale) will operate as a major mechanism of YRI improvements; high comorbidity will be a treatment modifier.*
> *Hypothesis 3(b)*: *Improvements in mental health and functioning due to YRI will lead to (mediate) greater employment outcomes and superior economic self-sufficiency over time.*
> *Hypothesis 3(c)*: *Homelessness, orphanhood, young parenthood, and high problems in emotion regulation comorbid with other mental health conditions will be moderators lessening the effectiveness of YRI.*

The table below outlines study primary outcomes, primary predictors, moderators and hypothesized mediators.

Primary Outcome Variables
Difficulties in Emotion Regulation (DERS)
Hopkins Symptom Checklist (HSCL)
Oxford Measure of Psychosocial Adjustment (OMPA)
Employment
Economic Self-Sufficiency
Primary Predictor Variables
Time
Group Membership (ENTR-only, YRI+ENTR, Control)
Possible Moderators
Gender
Age

Education Level
Literacy
Session Attendance
Hypothesized Mediators
Emotion regulation (DERS)
Oxford Measure of Psychosocial Adjustment (OMPA)
Hopkins Symptom Checklist (HSCL)
WHO Disability Assessment Scheduled (WHODAS)

Data Analysis Plan for Hypothesis 3a:

> Hypothesis 3(a): *Participants who are exposed to YRI will demonstrate greater reduction in mental health and functional problems than participants who do not receive the YRI intervention; emotion regulation (as measured using the Disturbances in Emotion Regulation Scale) will operate as a major mechanism of YRI improvements; high comorbidity will be a treatment modifier.*

Primary Outcomes DERS, HSCL, OMPA Externalizing, WHODAS

Primary Predictors: Time (baseline, Post-YRI, Post-ENTR, and 12 months and group membership (Control, ENTR, ENTR + YRI).

Possible Moderators: Subject gender, age, education level/literacy, session attendance

Strata Indicators: Subject ID and Cluster ID. With only three geographic/political districts, an insufficient number to constitute a level in multilevel modeling, the districts will be introduced at the top level (i.e., the clusters) as fixed effects by means of dummy variables to account for any district effects.

The goal of this analysis is to determine if there are differences among the three groups assigned at randomization (Control, ENTR, and ENTR+YRI). As described above, all analyses will include the aggregated baseline value of the outcome as a control at the cluster level, and furthermore, of any cluster-level potential treatment confounders shown to be imbalanced across the three groups. In particular we will be looking for group differences in the slopes of the primary outcomes over the lifetime of the study. Fixed effects will be time and group membership and their two-way interaction. Each subject will be measured three times—baseline, post-training, and at 12-month follow up—so we have three repeated measures for each subject, and each set of repeated measures is nested within study cluster. Subject- and cluster-specific slopes and intercepts will be modeled as random effects. We can test for the necessity of the subject and cluster specific slopes through likelihood ratio tests. Evidence of a group differential effects will be if the likelihood ratio test for the time by group interaction is significant at the 0.05 level. For pairwise differences we will examine the regression coefficients and base inference on Wald tests. We will estimate the model with Restricted Maximum Likelihood (REML, also MLR) for testing for random effect parameters and Full Maximum Likelihood (FML or MLF) for fixed effects. We will account for attrition by calculating inverse probability weights and/or multiple imputation. All linear modeling assumptions will be verified, both by evaluating underlying distributions and model residuals. Approaches to the violation of model assumptions or for non-continuous outcomes include use of response transformations, generalized linear mixed effect models or generalized estimating equations.

We now expand the above model to include possible effect modifiers gender, mental health status at baseline. To the model above we will include a three-way interaction between time, gender, and treatment group. Evidence of a moderating effect will be if likelihood ratio test for the three-way interaction term is significant at the 0.05 level.

As above, we will use cluster-level controls for aggregated levels of the outcome variable in all analyses, because even including the matching techniques employed, the total number of clusters is too small to assume that randomization will achieve balance on all variables. Furthermore, having tested for cluster-level imbalances in all cluster-level variables assessed at baseline, we will adjust models for any cluster-level confounders judged to have likely influence on the outcomes, for example, economic conditions, where significant imbalance was found. Similarly, individual-level confounders will be assessed in models and can be tested for their influence using multivariate generalized linear hypothesis testing to assess the influence of a group of variables or by means of log-likelihood test comparing nested models as described above.

We can use restricted fence methods as an example of variable selections.

Part II Emotion Regulation as the Mechanism Through which the YRI improvements operate

Primary Outcomes: OMPA internalizing, OMPA externalizing

Primary Predictor: Group Membership (Control, ENTR-only, YRI+ENTR

Hypothesized Mediator: Emotion Regulation (assessed by DERS)

Possible Moderators: Subject gender, age, education level/literacy, session attendance

We hypothesize that the YRI improvements may operate through changes in emotion regulation. This is a proposed mediation model where the primary predictor is group membership, the primary outcome is mental health at 12 months, and the mediator is emotion regulation as measured by the Difficulties in Emotion Regulation (DERS) at six weeks. We will test for the indirect effect of group membership on Mental Health Status using conditional process analysis (Hayes, 2013). The conditional process equations are (if gender modifies the indirect effect for example):

$$DERs(6\ weeks) = B_{0^*} + a^*X$$

$$MH\ (12\ months) = B_0 + cX + b_1{}^*DERS(6\ weeks) + b_2{}^*G + b_3{}^*G^*DERS.$$

Where X is group membership and G is gender. The coefficient a quantifies the effect of group membership on emotion regulation and b quantifies the effect emotion regulation on mental health independent of group membership. The product $p = a(b_1 + b_3{}^*G)$ is the *indirect effect* of group membership on Mental Health. We can test for the significance (null hypothesis $p = 0$) via bootstrapping. A significant (at 0.05) is indication that emotion regulation mediates the relationship between the YRI effects and mental health. For outcomes that are binary we will need to use the KHB method (Breen, Karlson, & Holm, 2013) to rescale the regression coefficients correctly. A strength of this analysis is that group membership, emotion dysregulation and Mental Health are measured at three different time points which is necessary for causal mediation (though not sufficient).

As an alternative we can test for mediation using Structural Equation Modeling (SEM). This approach is more flexible than conditional process analysis and may allow us to use full-information maximum likelihood (FIML) as opposed to multiple imputation. All causal models will be evaluated for the possible effects of unmeasured confounders (VanderWeele & Ding, 2017).

Data Analysis Plan for Hypothesis 3b:

> *Hypothesis 3(b): Improvements in mental health and functioning due to YRI will lead to (mediate) greater employment outcomes and superior economic self-sufficiency over time.*

Primary Outcome: Measures of employment and superior economic sufficiency.

Primary Predictor: Group membership.

Mediator: Mental Health as measured by the HSCL, OMPA; functioning as measured by WHODAS

Possible Moderators: Subject gender, age, education level/literacy, session attendance

Similar to Aim 3(a) we will test for mediation using condition process analysis the process equations are similar:

$$MH(12\ months) = B_{0^*} + a^*X$$

$$Employment = B_0 + B_0 + cX + b_1{}^*MH + b_2{}^*G + b_3{}^*G^*MH$$

The indirect effect of mental health on employment and economic self-sufficiency is quantified by the product $p = a(b_1 + b_3{}^*G)$. We will use a bootstrap test to access significance. Binary outcomes will need to be adjusted using KMB method. The size and direction of effects will be determined by the estimates of the regression coefficients.

Data Analysis Plan for Hypothesis 3c:

> *Hypothesis 3(c): Homelessness, orphanhood, young parenthood, and high problems in emotion regulation comorbid with other mental health conditions will be moderators lessening the effectiveness of YRI.*

Primary Outcome: Measures of employment and superior economic sufficiency

Primary Predictor: Group Membership

Youth FORWARD
Data Analysis Protocol
Feb. 6, 2020
Mediator: Mental Health

Possible Moderators: Subject gender, age, education level/literacy, session attendance, emotion regulation.

To the mediation model described in hypothesis 3b we will add terms to account for the moderating influences of gender, age, education level/literacy, session attendance, and emotion regulation. For purposes of exposition we will generically label the moderator as *W*. Our hypothesis is that moderators act on the influence of group membership and mental health. The conditional process equations in this case are (only one moderator was included for brevity):

$$MH(12\ months) = B_{0^*} + a^*X$$

$$Employment \quad = B_0 + B_0 + cX + b_1{}^*MH + b_2{}^*G + b_3{}^*G^*MH$$

In these equations the indirect effect of group membership on employment and economic self-sufficiency is $(a_1 + a_3{}^*W)^*b$. The value $a_3{}^*W{}^*b$ is sometimes called the index of moderated mediation. In the case of dichotomous W (gender, homelessness, young parenthood) this value reduces to $a_3{}^*b$ and is interpreted as the difference of the indirect effect of group membership on employment and economic self-sufficiency of the two *W* groups. Test for moderation can be directly from the coefficient a3. Alternative tactics include SEM as opposed to conditional process equations. KHB rescaling will be needed since moderators, mediators, and outcomes represent different variable types (binary, continuous etc.). In practice we can include all moderators in the model simultaneously.

We can also verify which path the candidate variables modify (we assumed the path from primary predictor to mediator). This can be done in the context of SEM models. We can fit models where the moderators act on the paths from mediator to outcome or both. Using fit indices such RMSEA we can compare model fits to determine which model best matches the data.

Missing Data: Using ITT principle we will obtain estimates of missing values of all study participants and at all time points. We will primarily use multiple imputations through chained equations. Missing data may exist at items within scales and we will use the Plumpton method (ref needed). Furthermore, where supported by our software, we can combine the use of imputations with attrition weighting. We will report both percentage of missing data and percentage of complete cases. For SEM models we can employ FIML to address missingness .

Multiple Comparisons: We will use False Discovery Rate (FDR) (ref needed) to adjust for multiple comparisons within hypothesis. This is be necessary since we will be comparing multiple models.

References

Breen, R., Karlson, K. B., & Holm, A. (2013). Total, direct, and indirect effects in logit and probit models. *Sociological Methods & Research*, *42*(2), 164-191.

Hayes, H. F. (2013) *Introduction to Mediation Moderation and Conditional Process Analysis*. The Guildford Press. New York, NY.

Higgins, M. J., Sävje, F., & Sekhon, J. S. (2016). Improving massive experiments with threshold blocking. *Proceedings of the National Academy of Sciences*, *113*(27), 7369-7376.

VanderWeele, T.J. and Ding, P. (2017). Sensitivity analysis in observational research: introducing the E-value. *Annals of Internal Medicine*, *167*, 268-274.