Andrew, Alison; Krutikova, Sonya

Research Report

Evaluation of an adolescent girl intervention in Rajasthan, India: Pre-analysis plan

IFS Report, No. R139

Provided in Cooperation with:
Institute for Fiscal Studies (IFS), London

Suggested Citation: Andrew, Alison; Krutikova, Sonya (2018) : Evaluation of an adolescent girl intervention in Rajasthan, India: Pre-analysis plan, IFS Report, No. R139, ISBN 978-1-911102-78-6, Institute for Fiscal Studies (IFS), London, https://doi.org/10.1920/re.ifs.2017.0139

This Version is available at:
https://hdl.handle.net/10419/201780

Institute for Fiscal Studies

Evaluation of an adolescent girl intervention in Rajasthan, India
Pre-analysis plan

Alison Andrew
Sonya Krutikova

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Evaluation of an Adolescent Girl Intervention in Rajasthan, India

Pre-Analysis Plan

December 2017

1. Background

This is a pre-analysis plan specifies analysis we plan to perform for the evaluation of PANKH: an intervention in Dhaulpur district, Rajasthan targeting marriage, education, mental health and socio-emotional skills of adolescent girls as well as gender attitudes of girls and their carers. Given the diverse range of outcomes covered the results of these analyses may be published separately. If we deviate from the methods set out in this document we will provide clear justifications and wherever possible will also publish per-plan analysis for comparison.

This study evaluates two models of delivering the PANKH intervention: (i) Girls only model and (ii) Girls + Community model. This is a cluster randomised controlled trial across 90 clusters. Stratifying by the three study blocks we allocated 30 clusters to each to: (i) control, (ii) girls only model, (iii) girls + community model. Full details of interventions and study design are available in the baseline report[1].

2. Endline Sample

At baseline we listed 12,281 eligible girls (12-17 unmarried or 12-19 married) living in 90 clusters. We selected 9,162 for the baseline survey and obtained complete baseline data for 7,574 girls[1]. The distribution of girls across treatment arms and marital status was as follows[2]:

	Control	**Girls only model**	**Girls + Community model**	**Total**
Never married	2027	1895	1956	**5878**
Married without gauna	26	38	29	**93**
Currently married	577	507	507	**1591**
Widowed/Divorced/Separated	3	7	2	**12**
Total	**2633**	**2447**	**2494**	**7574**

Table 1: Distribution of baseline sample across treatment arms and baseline marital status

[1] The baseline report reported this figure to be 7,577 but we subsequently discovered three corrupted entries.

[2] Note, married girls were oversampled so this does not accurately reflect the distribution of marital status in the population.

While PANKH targeted both married and unmarried girls, participation rates were very low amongst married girls despite extensive engagement activities targeting this group. Initial estimates of participation suggest only 7% of married girls attended any sessions compared with over half of unmarried girls.The study has insufficient power to detect any impact of PANKH on married girls with such low participation rates. We will therefore drop girls who were married at baseline from the quantitative evaluation and will not include them in sample for which we collect quantitative endline data. We will, however, include the 93 girls who were 'married without gauna' at baseline, meaning that their marriage had not been consummated and they were still living with their natal families. Following these changes, our target sample for endline data collection is distributed as follows:

	Control	**Girls only model**	**Girls + Community model**	**Total**
Never married	2027	1895	1956	**5878**
Married without gauna	26	38	29	**93**
Total	**2053**	**1933**	**1956**	**5971**

Table 2: Distribution of target endline sample across treatment arms and baseline marital status

Based on monitoring data collected during implementation, we anticipate that 10% of the sample will have moved from their baseline household (primarily for marriage). We plan to track and collect endline data from all girls who have moved within Dhaulpur district. If funds permit, we will extend this tracking to girls who have moved to other districts within Rajasthan.

3. Outcome Measures

We have selected primary and secondary outcomes on the basis of (i) PANKH's aims (outlined elsewhere[1]); and (ii) a power calculation exercise where we estimated the minimum detectable effects for possible outcomes based on the baseline data and compliance rates. The selected outcome measures (particularly primary outcomes) consist of those for which it will be possible to detect realistically sized treatment effects. Primary and secondary outcomes are listed in Tables 3 and 4 alongside information about each measure.

We will monitor the administration length of the questionnaire. If after the first two weeks of data collection the questionnaire is taking longer than required we will assess whether there is a need to cut one or more outcome measures.

Primary outcomes

Outcome	Measure	Unit of analysis	Details of measure	Baseline controls (in addition to 'core' controls). See baseline report for construction of scales[1].
1) Early marriage	a) Married	Girls	Binary indicator =1 if girl is currently married.	- Marriage fixed or engaged - Elders talking about marriage - Intended age of marriage (reported by caregiver): under 18, 18, over 18. - Gender attitude of caregiver.
	b) Married before age: 13, 14, 15, 16, 17, 18, 19	Girls	Binary indicators =1 if girl was married before age 13, 14, 15, 16, 17, 18, 19. Each indicator will be defined only for girls who are that age by the time of endline.	
2) Education	a) Currently in school	Girls	Binary indicator =1 if girl is currently attending school.	- Marriage fixed or engaged - Elders talking about marriage - Intended age of marriage (reported by caregiver): under 18, 18, over 18. - Gender attitude of caregiver.
	b) Dropped out from school before age: 13, 14, 15, 16, 17, 18, 19	Girls	Binary indicators =1 if girl dropped out of school before age 13, 14, 15, 16, 17, 18, 19. Each indicator will be defined only for girls who are that age by the time of endline.	
3) Mental health	Mental health factor	Girls	First principal factor from EFA of scales (3a) to (3c)	- Mental health (factor score)
	a) Depression – PHQ-9	Girls	Factor score(s) from PHQ-9 (9 items). The PHQ-9 is a brief version of the Patient Health Questionnaire (PHQ), an instrument for diagnosing depression and assessing the severity of depressive symptoms.[2] This brief version was originally validated for a US population but has since been validated for use in adults[3] and adolescent girls in India[4–6]. The adaptations/translations used in this study are based on those done on a sample of adolescent girls in Bihar[5,6].	- Mental health (factor score)
	b) Anxiety – GAD-7	Girls	Factor score(s) from GAD-7 (7 items). The GAD-7 is a brief self-report measure developed to identify cases of Generalised Anxiety Disorder (GAD). The measure was originally validated in US and German populations[7,8]. It has subsequently been adapted and	

			validated on a population of Indian adolescent girls[5,6]. The adaptations/translations used in this study are based on those done on a sample of adolescent girls in Bihar[5,6].	
	c) Rumination – RSS-10	Girls	Factor score(s) from RSS-10 (10 items) The Rumination Responses Scale (RSS) was developed to measure rumination, repetitive and passive thinking, in response to depressive moods[9]. The RSS-10 is a brief version where items with similar wordings to depression screeners have been removed[10]. This short version has been used in many contexts, including with adolescents[11,12]. We hypothesise that rumination could be important in understanding cognitive load and how girls are able to think about and take decisions. Therefore, we have translated, adapted and piloted the RSS-10 for use with adolescent girls in rural Rajasthan.	
4) Socio-emotional and non-cognitive skills	Socio-emotional and non-cognitive skills factor	Girls	First principal factor from EFA of (4a) to (4f)	- Self-efficacy (factor score) - Self-esteem (factor score) - Peer relations (factor score) - Socio-emotional skills (factor score)
	a) Self-efficacy – GSE-10	Girls	Factor score(s) from GSE-10 (10 items). The GSE-10 is a brief measure of self-efficacy, the belief in one's own ability to succeed in a given situations or task[13]. This measure has been used in over 25 countries[14] and was extensively piloted for use with Indian adolescents for the Young Lives study[15,16]. In addition, it was used in the baseline of this study where it was found to have good validity and reliability properties in our study population[1].	- Self-efficacy (factor score) - Self-esteem (factor score) - Peer relations (factor score) - Socio-emotional skills (factor score)
	b) Self-esteem – SDQ-GS	Girls	Factor score(s) from general-self sub-scale of SDQ (8 items). The Self Description Questionnaire (SDQ) measures self-concept in children and adolescents. The general-self scale, which drew on the Rosenberg self-esteem scale[17], was	

			added to the SDQ to capture general self-concept or self esteem[18,19]. The measure has been used in many populations and is well validated[20]. It has previously been used with adolescents in Andhra Pradesh, India as part of the Young Lives study[15,16] and was used in the baseline of this study where the data were found to have good validity properties[1].	
	c) Peer relations – SDQ-PR	Girls	Factor score(s) from peer relations sub-scale of the SDQ. The peer relations sub-scale of the SDQ measures children's and adolescents' relations with peers of their own age. This subscale has been well validated in many populations[20]. Amongst Indian adolescents, it performed well in Andhra Pradesh in the Young Lives study[15,16] and in the baseline to this study[1].	
	d) Resilience – CD-RISC-10	Girls	Factor score(s) from CD-RISC-10. (10 items) The CD-RISC-10 is a brief version of the Connor-Davidson Resilience Scale[21] (CD-RISC). The CD-RISC measures respondents' ability to cope with stress. The CD-RISC-10 both shortened the CD-RISC and created a unidimensional scale (i.e. a single factor was found in EFA)[22]. The CD-RISC-10 has been used on a population of adolescent girls in Bihar, India and these adaptions/translations form the basis for those used in this study[5,6].	
	e) Decision making – vigilance – DMQ	Girls	Factor score(s) from vigilance subscale of the DMQ. (6 items) The Melbourne Decision Making Questionnaire (DMQ) is a revised version of the older Flinders Decision Making Questionnaire[23]. It measures four decision-coping patterns. The vigilance sub-scale measures the respondent's tendency to clarify objectives, search for information, unbiasedly assimilate information and carefully evaluate alternatives. It is hypothesised that this is a pertinent	

			outcomes measure for the PANKH evaluation due to the curriculum encouraging girls to carefully consider the decisions that face them and think through their interests and possible options before coming to a decision.	
	f) Decision making – buck passing – DMQ	Girls	Factor score(s) from buck-passing sub-scale of the DMQ. (6 items) The buck-passing sub-scale measures respondents' tendency to avoid making decisions and to leave making decisions to others[23]. It is hypothesised that this is a pertinent outcomes measure for the PANKH evaluation due to the curriculum encouraging girls to take an active role is decisions involving their own welfare and giving them skills to negotiate with others around these decisions.	
5) Gender attitudes	Gender attitudes factor	Girls + Carers	First principal factor from EFA of (5a) and (5b)	- Gender attitude of girls (factor score) - Gender attitude of carers (factor score)
	a) Gender attitudes of girls	Girls	Factor score(s) from gender attitudes scale (14 items) This gender attitudes scale is partially based on the Gender Equitable Men Scale (GEMS) which has been adapted and expanded by ICRW for various research projects. The baseline of this study used 29 items. Exploratory factor analysis gave two factors, the first capturing strongly patriarchal gender attitudes including advocating violence towards women who violate such norms[1]. For evaluating the impact of PANKH, all 14 items which loaded onto the first (rotated) factor with a loading of >0.3 have been kept to create a brief unidimensional measure of patriarchal gender attitudes[3].	- Gender attitude of girls (factor score) - Gender attitude of carers (factor score)

[3] 1. Boys should be given more privilege as compared to the girls, 2. Women/girls should work only if there are monetary needs in their family, 3. Only bad girls make male friends, 4. A man should have the final say in all family matters, 5. Men should be more educated than their wives, 6. Boys are naturally better than girls in studies, 7. A daughter deserves to be beaten if she does not obey her parents,

	b) Gender attitudes of caregivers – GEMS	Caregivers	Factor score(s) from gender attitudes scale (14 items) Gender attitudes of caregivers are measured using the same shortened version of the baseline instrument as described above.	

Table 3: Primary Outcomes

8. Girls should be married early to protect them from sexual harassment, 9. Girls should be married early to ease family's financial burden, 10. Instead of spending money on a girl's education, it should be saved for her dowry, 11. If a girl is a victim of some sexual abuse, it is the fault of the girl, 12. A woman should tolerate violence in order to keep her family together, 13. There are times when a woman deserves to be beaten, 14. Girls who are highly educated indulge in improper behaviour

Secondary outcomes

Outcome	Measure	Unit of analysis	Details of measure	Baseline controls (in addition to 'core' controls). See baseline report for construction of scales[1].
1) Attitude towards school	School attitudes scale (9 items)	Girls	Factor score(s) from school attitudes scale (9 items). Scale created by research team though combining previous statements used in other surveys researchers were involved with and new statement. Statements comprise: 1) I enjoy school 2) I am motivated to work hard at school 3) I am bored in school 4) What I learn at school will be useful for my future 5) I want to quit school 6) I learn lots of new things at school 7) I look forward to going to school 8) My teachers at school want me to do well 9) I feel my studies have no meaning	
2) Knowledge of sexual and reproductive health	Knowledge of sexual and reproductive health factor	Girls (aged 15+)	First principal factor from EFA of scales (2a) and (2b)	Menstruation knowledge scale Contraceptive knowledge scale
	a) Puberty and menstruation Knowledge scale	Girls	Score predicted using IRT on puberty and menstruation knowledge scale (19 items). This is a series of 19 multiple choice questions and true/false statements addressing puberty and menstruation. Each will be scored (=1 if correct, 0 otherwise) and a score constructed through an IRT analysis these indicator variables.	Menstruation knowledge scale
	b) Contraception and sexual health knowledge scale	Girls (aged 15+)	Score predicted using IRT on contraception and sexual health knowledge scale (19 items). This is a series of 19 multiple choice questions and true/false statements addressing contraception and menstruation. Each will be scored (=1 if correct, 0 otherwise) and a score constructed through an IRT analysis these indicator variables.	Menstruation knowledge scale Contraceptive knowledge scale

3) Attitudes and responses to violence	Attitudes and responses to violence factor	Girls	First principal factor from EFA of scales (3a) to (e) below.	Gender attitudes scale
	a) 'Victim blaming' tendencies for violence against women and girls	Girls	Score predicted using IRT on 'victim blaming' scale (6 items). This is a series of 6 indicator variables created from girls' responses to 6 vignettes describing situations of violence against women and girls. For each vignette, the corresponding indicator is scored =1 if the girl answers that the female victim of violence in the vignette is either wholly or partially to blame for the incident.	Gender attitudes scale
	b) 'Perpetrator blaming' tendencies for violence against women and girls	Girls	Score predicted using IRT on 'perpetrator blaming' scale (6 items). This is a series of 6 indicator variables created from girls' responses to 6 vignettes describing situations of violence against women and girls. For each vignette, the corresponding indicator is scored =1 if the girl answers that the perpetrator of violence in the vignette is either wholly or partially to blame for the incident.	Gender attitudes scale
	c) 'Avoidance behaviours' in response to violence against women and girls	Girls	Score predicted using IRT on 'avoidance behaviours' scale (6 items). This is a series of 6 indicator variables created from girls' responses to 6 vignettes describing situations of violence against women and girls. For each vignette, the corresponding indicator is scored =1 if the girl answers that the female victim of violence should avoid the situation in which the violence happened in future.	Gender attitudes scale
	d) 'Retaliation behaviours' in response to violence against women and girls	Girls	Score predicted using IRT on 'avoidance behaviours' scale (6 items). This is a series of 6 indicator variables created from girls' responses to 6 vignettes describing situations of violence against women and girls. For each vignette, the corresponding indicator is scored =1 if the girl answers that the female victim of violence should seek retaliation against the perpetrator.	Gender attitudes scale
	e) 'Reporting behaviours' in response to violence	Girls	Score predicted using IRT on 'reporting behaviours' scale (6 items).	Gender attitudes scale

	against women and girls		This is a series of 6 indicator variables created from girls' responses to 6 vignettes describing situations of violence against women and girls. For each vignette, the corresponding indicator is scored =1 if the girl answers that the female victim of violence should seek report the perpetrator.	
	Attitudes and responses to violence factor	Carers	First principal factor from EFA of scales (3f) to (j) below.	Gender attitudes scale (carers)
	f) 'Victim blaming' tendencies for violence against women and girls	Carers	Score predicted using IRT on 'victim blaming' scale (3 items). This is a series of 3 indicator variables created from carers' responses to 3 vignettes describing situations of violence against women and girls. For each vignette, the corresponding indicator is scored =1 if the carer answers that the female victim of violence in the vignette is either wholly or partially to blame for the incident.	Gender attitudes scale (carers)
	g) 'Perpetrator blaming' tendencies for violence against women and girls	Carers	Score predicted using IRT on 'perpetrator blaming' scale (3 items). This is a series of 3 indicator variables created from carers' responses to 3 vignettes describing situations of violence against women and girls. For each vignette, the corresponding indicator is scored =1 if the carer answers that the perpetrator of violence in the vignette is either wholly or partially to blame for the incident.	Gender attitudes scale (carers)
	h) 'Avoidance behaviours' in response to violence against women and girls	Carers	Score predicted using IRT on 'avoidance behaviours' scale (3 items). This is a series of 3 indicator variables created from carers' responses to 3 vignettes describing situations of violence against women and girls. For each vignette, the corresponding indicator is scored =1 if the carer answers that the female victim of violence should avoid the situation in which the violence happened in future.	Gender attitudes scale (carers)
	i) 'Retaliation behaviours' in response to violence against	Carers	Score predicted using IRT on 'avoidance behaviours' scale (3 items). This is a series of 3 indicator variables created from carers'	Gender attitudes scale (carers)

	women and girls		responses to 3 vignettes describing situations of violence against women and girls. For each vignette, the corresponding indicator is scored =1 if the carer answers that the female victim of violence should seek retaliation against the perpetrator.	
	j) 'Reporting behaviours' in response to violence against women and girls	Carers	Score predicted using IRT on 'reporting behaviours' scale (3 items). This is a series of 3 indicator variables created from carers' responses to 3 vignettes describing situations of violence against women and girls. For each vignette, the corresponding indicator is scored =1 if the carer answers that the female victim of violence should seek report the perpetrator.	Gender attitudes scale (carers)
4) Sports	a) Attitude to and enjoyment of sports – sports attitudes scale (10 items)	Girls	Factor score(s) from sports attitudes scale (10 items). This scale is an expanded version of a sports attitudes scale used at baseline. 1) I can play sports with the other girls in my village. 2) I can make friends by playing sports 3) I can be healthier by playing sports. 4) I can encourage other girls to take up sports in my community. 5) I can ask for play time after I complete household chores. 6) I can be comfortable with my body while playing sport 7) I can compete in sports just as well as boys. 8) I enjoy playing sports 9) I feel safe when I play sports 10) When I play sports I feel I have time for myself	Sports attitudes
	b) Participation in sports	Girls	Time spent playing sport in a typical week (minutes)	Frequency of playing sports.
5) Restrictions during menstruation	Restrictions during menstruation	Girls	Score predicted using IRT on restrictions during menstruation scale (7 items). This scale was used at baseline and comprises of 7 items, =1 if the girl does each of the following activities during menstruation: 1) Attend religious function 2) Cook 3) Touch stored food 4) Sleep in your usual bed/ place	Restrictions during menstruation

			5) Touch family members 6) Play outside/ see friends outside house 7) Visit relatives	

Table 4: Secondary outcomes.

4. Empirical Strategy

4.1. Attrition Analysis

We will test for differential attrition by treatment status and balance of our endline sample to ensure the randomisation has not been compromised by different rates of attrition across the different treatment arms.

We will test whether attrition, A_{ij}, is significantly related to treatment, T_j^G and T_j^I, status through the following logistic regression:

$$A_{ij} = \beta_G T_j^G + \beta_I T_j^I + \gamma X_{ij} + \varepsilon_{ij} \quad (1)$$

We will estimate the parameters both with and without controlling for key household and child baseline characteristics X_{ij} such as age, household wealth, caste, education etc. We will test the null hypothesis that $\beta_G = \beta_I = 0$, i.e. there is no differential attrition by treatment status, accounting for clustering of errors (ε_{ij}) in our analysis. If we reject this hypothesis this implies that there is differential attrition between treatment groups.

If we do find evidence of differential attrition across treatment groups we will explicitly model the attrition (based on observable characteristics at baseline) and report estimates that are corrected for it, as well as bounds on the estimates where necessary.

We will assess the baseline balance of the endline sample by assessing the difference in means between treatment arms across baseline characteristics, accounting for clustering of errors and multiple hypothesis testing. We will assess baseline balance across: caste, wealth of household, age of girl, whether in school, highest grade of school, gender attitudes, gender attitudes of carer, mental health, socio-emotional skills, self-esteem, self-efficacy and peer-relations. All measures are outlined in the baseline report and all scales will be constructed using factor analysis as set out in this document[1].

4.2. Empirical specification

Our main analysis will be an intent-to-treat analysis of the impact of eligibility for the PANKH interventions. For each outcome of interest we will estimate the following regression:

$$y_{ij} = \alpha + \beta_G T_j^G + \beta_I T_j^I + \gamma X_{ij} + \varepsilon_{ij} \quad (2)$$

where y_{ij} is the outcome of interest for girl (or household) i, in village j, T_j^G is a dummy variable equal to one if village j was allocated to the girls' only treatment group and equal to zero otherwise and T_j^C is a dummy variable equal to one if village j was allocated to the girls+community treatment group and equal to zero otherwise. To increase the precision of our estimates we will control for key baseline chacteristics, X_{ij}. The variables contained in X_{ij} for each outcome are specified below (Section 4.4). In all inference we will allow for clustering of the random error term ε_{ij} at the unit of randomization (cluster level).

In this regression framework the most interesting parameters are β_G and β_C, our estimates of the impact of being eligible for the girls' only programme and the integrated programme respectively.

The size and significance of these parameters will inform on the impact of the interventions on the outcomes of interest and the degree of uncertainty associated with each estimate. Therefore, for each outcome of interest we will perform the following hypothesis tests: (i) $H_0: \beta_G = 0$, (ii) $H_0: \beta_C = 0$ and (iii) $H_0: \beta_G = \beta_C$.

In the interests of power, we will also estimate the impact of being eligible for either intervention compared to the control group through the following specification:

$$y_{ij} = \alpha + \beta_{GC} T_j^{GC} + \gamma X_{ij} + \varepsilon_{ij} \quad (3)$$

where T_j^{GC} is a dummy variable equal to one if village j was allocated to either the girls' only treatment group or the girls + community treatment group. Here we will test the hypothesis: $H_0: \beta_{GC} = 0$

For all continuous outcome measures we will estimate (2) and (3) using Ordinary Least Squares (OLS) while in the case of binary outcome variables we will use a logistic regression model.

4.3. Heterogeneous Effects

We assess heterogeneous treatment effects on primary and secondary outcomes (Tables 3 and 4) over various dimensions using an interaction between H_{ij}, a dummy variable either equal to 1 or 0 depending on the characteristic we are testing for heterogeneous effects over, and the treatment indicators.

$$y_{ij} = \alpha + \beta_G T_j^G + \beta_I T_j^I + \alpha_G T_j^G H_{ij} + \alpha_I T_j^I H_{ij} + \delta H_{ij} + \gamma X_{ij} + \varepsilon_{ij} \quad (4)$$

In this specification our estimated effect of the intervention on the group for which $H_{ij} = 0$ will be β_G or β_I and on the group for which $H_{ij} = 1$ it will be $\beta_G + \alpha_G$ or $\beta_I + \alpha_I$. Therefore, we will test the null hypothesis of homogenous treatment effects over characteristic H_{ij} by testing the null hypotheses $\alpha_G = 0$ and $\alpha_I = 0$.

We will assess heterogeneity of effects by:

- Age
- Wealth (asset ownership at baseline)
- Caste
- Baseline measure of outcome in question

4.4. Control Variables

Depending on whether the outcome in question is measured at the girl or caregiver level we will control for the following characteristics:

(i) Girl level core controls:
- Full set of interactions between: dummy variables for age in years and dummy variable indicating whether girl was in school at baseline
- Caste: Dominant Caste, OBC, SC, ST, other

- Baseline wealth index
- Mother's education in years

(ii) Caregiver core controls:
- Age in years
- Age in years squared
- Education in years
- Caste: Dominant Caste, OBC, SC, ST, other
- Baseline wealth index

In addition to these core controls we will control for baseline measures that are closely related to the outcome of interest. In Tables 3 and 4 we specify the additional baseline controls we will use for each outcome category.

In the case of missing baseline controls we will impute the missing covariate value with the average (mean for continuous controls, median for discrete) of the non-missing observations and this imputation will be accounted for with a dummy variable equal to one for imputed observations.

4.5. Multiple Hypothesis Testing

We plan to deal with multiple hypothesis testing in two ways.

Firstly, within each outcome (e.g. socio-emotional skills) we plan to reduce the dimensionality of continuous measures into a single or multiple index/indices using factor analysis. We will then test whether the estimated effect on this/these aggregated index is significantly different from zero. We will use exploratory factor analysis using the iterated principal factor methodology. We will keep and rotate as many factors as there are with eigenvalues greater than 1 and assess the effect of the intervention(s) on this/these summary index/indices.

Secondly, when testing whether the treatment(s) have an impact on each outcome measure we will adjust the p-values for multiple hypothesis testing within each outcome category (e.g. marriage, socio-emotional skills etc). We will use the Romano-Wolf step-down procedure for this adjustment[24,25].

The first two columns in Tables 3 and 4 show measures (column 2) will be considered to be within the same outcome (column 1) when creating single aggregate indices and carrying out the multiple hypothesis testing adjustments to p-values.

4.6. Scoring of Scales

To score all scales with Likert type response options (e.g. Strongly Agree, Agree,...) we will use exploratory factor analysis (iterated principal factor) to construct factor score(s) for each scale. We will run an exploratory factor analysis of all items in the scale. We will then retain and rotate all factors with an eigenvalue greater than 1 using a varimax rotation and assess the estimated factor loadings. We will drop any items that load onto the (both) retained factor(s) with a loading less than 0.3 in absolute magnitude and any items that load in the opposite direction than anticipated. From baseline analysis of scales we anticipate this to be the case for very few items. If we drop any items we will then rerun the factor analysis with all remaining items. We will then

estimate the underlying factor(s) from the factor loadings. Analysis will be performed using STATA's factor command using the 'ipf' option.

To score all scales with binary response options (e.g. yes/no, true/not true) we will use an item response theory (IRT) two parameter model. We will use a latent factor model to estimate the difficulty and discrimination parameters associated with each item and predict the underlying score using an empirical bayes methodology. We will drop items whose discrimination parameters are not significantly different from zero or whose discrimination parameters are negative.

4.7. Modifications/piloting of scales

References and justifications for each scale used are summarized in Tables 3 and 4. Wherever possible we use scales that were used in our population at baseline or amongst adolescent girls in culturally and socio-economically similar areas of India[5,6]. Adaptations, where necessary, were made to all scales on the basis of piloting to ensure functional equivalence of items. In cases where the original intent of the item was not being well understood, concepts were defined or examples given. Wherever it made sense, all scales were modified to use a 4 point Likert scale (Strongly Agree, Agree, Disagree, Strongly Disagree).

References

1 Pranita Achyut, Andrew A, Das M, *et al.* Promoting Adolescent Engagement , Knowledge and Health Evaluation of PAnKH : an adolescent girl intervention in Rajasthan , India Baseline Report. 2016.

2 Kroenke K, Spitzer RL, Williams JBW. The PHQ-9. *J Gen Intern Med* 2001; **16**: 606–13.

3 Patel V, Araya R, Chowdhary N, *et al.* Detecting common mental disorders in primary care in India: a comparison of five screening questionnaires. *Psychol Med* 2008; **38**: 221–8.

4 Ganguly S, Samanta M, Roy P, Chatterjee S, Kaplan DW, Basu B. Patient health questionnaire-9 as an effective tool for screening of depression among indian adolescents. *J Adolesc Heal* 2013; **52**: 546–51.

5 Leventhal KS, DeMaria LM, Gillham J, Andrew G, Peabody JW, Leventhal S. Fostering emotional, social, physical and educational wellbeing in rural India: the methods of a multi-arm randomized controlled trial of Girls First. *Trials* 2015; **16**: 481.

6 Leventhal KS, Gillham J, DeMaria L, Andrew G, Peabody J, Leventhal S. Building psychosocial assets and wellbeing among adolescent girls: A randomized controlled trial. *J Adolesc* 2015; **45**: 284–95.

7 Spitzer RL, Kroenke K, Williams JBW, Löwe B. A Brief Measure for Assessing Generalized Anxiety Disorder. *Arch Intern Med* 2006; **166**: 1092.

8 Löwe B, Decker O, Müller S, *et al.* Validation and Standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the General Population. *Med Care* 2008; **46**: 266–74.

9 Nolen-Hoeksema S, Morrow J. A prospective study of depression and posttraumatic stress symptoms after a natural disaster: the 1989 Loma Prieta Earthquake. *J Pers Soc Psychol* 1991; **61**: 115.

10 Treynor W, Gonzalez R, Nolen-Hoeksema S. Rumination reconsidered: A psychometric analysis. *Cognit Ther Res* 2003; **27**: 247–59.

11 Ruiz FJ, Sierra MA, Montero KB, *et al.* Psychometric properties and factor structure of the ruminative responses scale-short form in Colombia. *Int J Psychol Psychol Ther* 2017; **17**: 199–208.

12 Erdur-Bakera Ö, Bugaya A. The short version of ruminative response scale: Reliability, validity and its relation to psychological symptoms. *Procedia - Soc Behav Sci* 2010; **5**: 2178–81.

13 Schwarzer R, Jerusalem M. Generalized Self-Efficacy scale. In: Measures in health psychology: A user's portfolio. Causal and control beliefs. Windsor, England: NFER-NELSON, 1995: 35–7.

14 Scholz U, Doña BG, Sud S, Schwarzer R. Is general self-efficacy a universal construct? Psychometric findings from 25 countries. *Eur J Psychol Assess* 2002; **18**: 242–51.

15 Pells K, Portela MJO, Revollo PE. Experiences of peer bullying among adolescents and associated effects on young adult outcomes: longitudinal evidence from Ethiopia, India, Peru and Viet Nam. *Discuss Pap UNICEF Innocenti Res Cent* 2016; **3**.

16 Singh A, Espinoza Revollo P. Teenage Marriage, Fertility and Wellbeing: Panel Evidence from India. 2016.

17 Rosenberg M. Society and the adolescent self-image. Princeton university press Princeton, NJ, 1965.

18 Marsh HW, Smith ID, Barnes J. Multidimensional self-concepts: Relations with sex and academic achievement. *J Educ Psychol* 1985; **77**: 581–96.

19 Marsh HW, O'Neill R. Self Description Questionnaire III: The construct validity of multidimensional self-concept ratings by late adolescents. *J Educ Meas* 1984; **21**: 153–74.

20 Byrne BM. Measures of self-concept for adolescents. American Psychological Association, 1996.

21 Connor KM, Davidson JRT. Development of a new Resilience scale: The Connor-Davidson Resilience scale (CD-RISC). *Depress Anxiety* 2003; **18**: 76–82.

22 Campbell-Sills L, Stein MB. Psychometric analysis and refinement of the connor–davidson resilience scale (CD-RISC): Validation of a 10-item measure of resilience. *J Trauma Stress* 2007; **20**: 1019–28.

23 Mann L, Burnett P, Radford M, Ford S. The Melbourne Decision Making Questionnaire: An instrument for measuring patterns for coping with decisional conflict. *J Behav Decis Mak* 1997; **10**: 1–19.

24 Romano JP, Wolf M. Stepwise Multiple Testing as Formalized Data Snooping. *Econometrica* 2005; **73**: 1237–82.

25 Romano JP, Wolf M. Exact and Approximate Stepdown Methods for Multiple Hypothesis Testing. *J Am Stat Assoc* 2005; **100**: 94–108.