Assessing the effect of the Samata intervention on secondary school completion and child marriage among adolescent girls from marginalized communities in north Karnataka, India: Results from a cluster-randomised trial

Statistical Analysis plan

1. Removal of programme exposure data from the endline dataset by Prakash Javalkar
2. Transfer of endline dataset to the Trial Statistician (Ravi Prakash)
3. Cleaning of data to include consistency checks (due to time constraints, data cleaning will be limited to the characteristics of girls and primary and secondary outcome variables listed in the dummy Tables below).
4. Append Trial start and endline datasets. (We realise that 'Trial start' would usually be referred to as baseline. However as cohort 2 Trial start data collection took place 3 months after the intervention began, it is misleading to call it 'baseline'. Therefore, in this document we will refer to the cohort 1 and cohort 2 datasets collected in March and November 2014 as 'Trial start', although we may use different terminology agreed by the team for the Trial paper).
5. Coding of the characteristics of girls and primary and secondary outcome variables needed for the Trial paper (listed in the dummy Tables below) and checking for missing data. We will use the same coding as was developed for the Trial start datasets, except for those outcomes which weren't measured at the start of the Trial. We will include a binary variable for Strata.
6. Creation of Figure 2 – The Consort diagram. This has already been partially presented in the Samata protocol paper. This will be added to, to include data at Trial start and Endline, by Trial arm. For the workshop in Mysore, the Trial arms data will be blinded.
7. Randomly assign the 80 village clusters to intervention and control villages (40 in each arm).
   Creation of Table 1. Present N and % at Trial Start for the combined dataset and for cohort 1 and cohort 2 separately (see Table 1 below). We will present individual-level data for section A in Table 1 below. E.g. tab arm agegp if round==0, col; tab arm agegp if cohort==0 & round==0, col; tab arm agegp if cohort==1 & round==0, col.
   We will present cluster-level summaries for sections B and C as we are really interested in whether there is baseline imbalance between village clusters. E.g. collapse arm agegp, by cluster if round==0; collapse arm agegp, by cluster if round==0 & cohort==0; collapse arm agegp, by cluster if round==0 & cohort==1. Use tabout command to generate dummy Tables for Table 1. For the initial analysis, we will use unweighted data. However if there is a big difference in the size of the clusters, after the Mysore workshop, we will re-analyse and present weighted cluster results.
8. Assess extent of missing data (including loss-to-follow-up and refusal to participate at endline) across outcomes and explanatory variables documenting proportions. Collect and report information about reasons for missing values. Discuss options for sensitivity analyses.
9. Creation of Table 2. Cluster level summaries (mean of the cluster-level means by Trial arm) for each primary and secondary outcomes.
   To generate cluster level summary by Trial arm: collapse outcome1 arm, by (cluster)
   To generate risk difference and CI: ttest oucome1, by (arm) (see page 196-7 Hayes and Moulton book).

To generate intracluster correlation: using cltest command

10. Creation of Table 3. Intervention effect at the individual level. We will use a multi-level model, adjusting for clustering and for strata. (see p212 Hayes and Moulton book). For binary outcomes like ours, this is known as logistic regression with random effects.
    To generate N and percentages: tab arm outcome1 if round==1, col; tab arm outcome1 if cohort==0 & round==1, col; tab arm outcome1 if cohort==1 & round==1, col
    To generate OR for the basic model which accounts for clustering at the village level: xi:xtlogit outcome1 i.arm, or re i(cluster)
    To generate OR adjusted for Trial start imbalances: xi:xtlogit outcome1 i.arm i.strata i.variable1 i.variable2 ……., or re i(cluster). If we are adjusting for any baseline imbalance, we will use cluster-level baseline data for that variable.
    Check the reliability of the model using quadchk command.
    Where there are composite measures, results of the individual measures will be presented at the workshop in Mysore.
11. Once do files are running well, independent statistician (Rajaram) to send Trial statistician (Ravi), blinded coding for clusters (intervention / control). Trial statistician to complete blinded Figure 2 and steps 8-10 above on the blinded data. Following output from step 8 using blinded Trial arms (Table 1), we will decide which Trial Start characteristics need to be adjusted for at Endline (step 10, Table 3).
12. Following presentation and discussion of blinded results at the Mysore workshop (4-8th dec 2017), results will be unblinded.
13. Discussion and additional analyses will be required to understand missing data, loss-to-follow-up and refusal to participate at endline data and the effects we think these have had on the Trial. This will be done in two ways: i) although imperfect, where the team has it, data on Trial outcomes for girls lost-to-follow-up / refused to participate will be presented in a separate Table – potentially as an appendix - to describe missing data for the Trial arms. ii) Imputation may be used for missing data >25% and <75%, although further discussions are needed before this is decided upon.

Dummy tables and figures

Figure 1. Trial Timeline (Intervention and Data Collection Time Lines)

Figure 2. Trial Profile (CONSORT diagram)

Table 1. Assessing cluster level imbalance in characteristics of girls and trial outcomes across intervention and control arms at the beginning of the trial

	Combined				Cohort 1			Cohort 2		
	N %				N %			N %		
	Overall	Intervention	Control	Mean difference	Overall	Intervention	Control	Overall	Intervention	Control
Total										
A. Socio-demographic Characteristics*										
Cohort										
Village Type (Feeder vs. main)										
District										
Median age of adolescent girls (years)										
Caste (ST/SC)										
Menarche (N/Y)										
Orphan hood (No/One or Both Parents)										
Household wealth quintile (Poorest – Richest)										
Non-literate household heads (Literate/Illiterate)										

B. Sibling and school level characteristics ** Recent eve teasing (N/Y) Girls having a sister who dropped out of school before age 16 (N/Y) Girls having a sister who married <18 years (N/Y) Girls who report poor learning environment in current / last school (N/Y) Girls who report poor school infrastructure in current / last school (N/Y) Girls who report harassing/bullying environment in current / last school (N/Y)										
C. Schooling, marriage and sexual debut at Trial start** Proportion of girls who dropped out of school Proportion of girls who are married Proportion of girls who have sexual debut Proportion of girls married and co-habiting with husband										

***Individual level data** ****Cluster-level summaries (mean of the mean)**

Table 2. Cluster-level summaries of the outcomes at endline

	Cluster Level Summary		Intra-cluster correlation	Risk Difference (95% CI)
	Control	Intervention		
	N (%)	N (%)		
Primary outcomes				
Proportion of girls who complete secondary school [sit 10th standard exam]				
Proportion of girls who are married [by Trial end line]				
Secondary outcomes				
Proportion of girls who start secondary school [enter into 8th standard]				
Proportion of girls who pass secondary school final year exams[pass 10th standard exam]				
Proportion of girls who have sexual debut [by Trial end line]				
Proportion of girls married and co-habiting with husband [by Trial end line]				

Table 3. Effects of the intervention on outcomes at endline

	Summary Statistics		Intervention effect*			
	Control	Intervention	Basic Model		Adjusted model	
	N (%)	N (%)	OR (95% CI)	P value	AOR (95% CI)	P value
Primary outcomes						
Proportion of girls who complete secondary school [sit 10th standard exam]						
Proportion of girls who are married [by Trial end line]						
Secondary outcomes						
Proportion of girls who start secondary school [enter into 8th standard]						
Proportion of girls who pass secondary school final year exams [pass 10th standard exam]						
Proportion of girls who have sexual debut [by Trial end line]						
Proportion of girls married and co-habiting with husband [by Trial end line]						
*Individual level logistic regression with random effects						