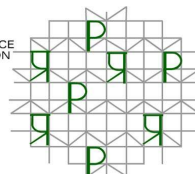# Promoting Adolescent Engagement, Knowledge and Health (PAnKH) in Rajasthan, India

Programme Impact Evaluation

Institute for Fiscal Studies

ICRW International Center for Research on Women
where insight and action connect

PROFESSIONAL ASSISTANCE FOR DEVELOPMENT ACTION
Pradan

Promoting Adolescent Engagement, Knowledge and Health (PAnKH) in Rajasthan, India

Programme Impact Evaluation

Institute for Fiscal Studies (London, UK)

Alison Andrew, Sonya Krutikova, Gabriela Smarrelli*

International Center for Research on Women (Delhi, India)

Hemlata Verma, Abhishek Gautam, Ravi Verma, Madhumita Das*, Pranita Achyut*, Ronak Soni*

Professional Assistance for Development Action (Dhaulpur, India)

Sanjay Sharma

Address correspondence to:

Alison Andrew, alison_a@ifs.org.uk

Hemlata Verma, hverma@icrw.org

*Previous affiliation

December 2018

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

1 INTRODUCTION

Women in developing countries continue to be disempowered facing multiple constraints which prevent them from investing in their human capital and breaking the cycle of dependence on men. These include high youth unemployment, low wages, as well as early marriage and child-bearing (World Bank, 2007; Jayachandran, 2015). India is a particularly salient case. Norms and attitudes centred on the primacy of men as decision makers and on women as holding a family's honour create environments where it is difficult for young women to pursue their education, where many marry early and where they are unequipped with the skills and knowledge needed to make choices that are optimal for their future.

There is some encouraging evidence suggesting that interventions which jump-start women's human capital through building up different skills may have the potential to set them on a better trajectory (e.g. Case & Paxon, 2013; Adhvaryu et al, 2016). From a developmental perspective, adolescence is an opportune time for such interventions as this is a period of profound transformations in the brain, particularly in the development of higher cognitive functions and socio-emotional skills which are critical for long-term success (Heckman & Rubinstein, 2001; Heckman et al, 2006; Fuhrmann et al, 2015) offering a 'window of opportunity' during which appropriate interventions may have lifelong impacts (Eldreth et al, 2013). Combined these pieces of evidence suggest that interventions targeting human capital during adolescence may be a promising way forward in achieving better outcomes for women in the long-run.

Consistently with this the Indian government has shown a growing interest in adolescent groups with a number of policies and programmes initiated over the last two decades. The focus of these has gradually broadened from that restricted to girls' physical health and school attendance, to include life-skills, empowerment and knowledge of sexual and reproductive health. The aim is to target the barriers that adolescent girls face to securing better economic and psychosocial outcomes[1].

Much of the evidence that exists on the potential of interventions in adolescence to improve key outcomes such as schooling and marriage focuses on programmes that either relax financial resource constraints such as cash transfers (Baird et al, 2011, other), focus on specific vocational skills training only (e.g. Field et al, 2010), or are very "bundled" and include life-skills training and empowerment alongside a number of other components such as vocational skills training, cash transfers, microfinance initiatives etc. (e.g. Buchmann et al, 2017; Adoho et al 2015). There is a repeated finding that cash transfers in various forms can be effective at improving girls' schooling and marriage outcomes (Baird et al, 2011; Duflo et al, 2015; Duflo et al, 2017; Buchmann et al, 2017).

---

[1] These include programmes such as *Beti Bachao Beti Padao* (BBBP) and *Rashtriya Kishor Swasthya Karyakarm* (RKSK). Detailed descriptions can be found in the companion *Implementation and Cost Effectiveness Analysis* report.

The most relevant studies for informing the current approach of the government of India, however, are those looking at impacts of stand-alone life-skills and empowerment programmes. There are not many of these to begin with and most study impacts on mental health and socio-emotional well-being (Barry et al, 2013; Leventhal et al, 2015). Only a handful study impacts on key life outcomes such as education and marriage and the evidence that exists for South Asia is mixed: one of the few rigorous trials of a life-skills programme for adolescent girls in Bangladesh showed no impacts on marriage and childbearing but some indication of positive schooling effects (Buchmann et al, 2017). From a theoretical perspective it is also unclear how such programmes might impact targeted outcomes including marriage and education in contexts where girls are particularly disempowered and have even less say on key decisions regarding, among others, their schooling, marriage and family planning than do girls in many other developing country settings.

In addition to inconclusive evidence on the feasibility of changing schooling and marriage outcomes through life-skills and empowerment interventions, especially in a context such as Rajasthan, it is also unclear whether the implementation approach favoured by the government is effective. Key schemes are highly reliant on 'peer education' models whereby the responsibility is placed on adolescents themselves to organise and run group sessions to impart the relevant skills, knowledge and empowerment. The potential benefits of such an approach are (1) tapping into adolescents' heightened sensitivity to peer influence (e.g. Dahl et al, 2018); and (2) cost-saving. However, it remains a question whether peer educators can master the methods and materials sufficiently well to provide the quality of input needed to make a difference, especially given that many of the programmes allow for only very low intensity of training.

In sum, much of the existing evidence gives reason to be optimistic about the current approach of the Indian government to programming for adolescents – developmentally this is a promising stage for intervention and there is evidence that it might be possible to "jump-start" women's human capital through adolescent intervention - but there are important gaps in the knowledge base on whether scalable models of life skills interventions can have such a jump-starting effect. In order to fill these gaps researchers at ICRW, PRADAN and IFS teamed up to design and evaluate a programme – *PanKH* – which, if found to be effective, could serve as a blueprint for scalable and cost-effective programmes targeting adolescent girls' education, early marriage, life skills and physical and mental health and inform the design and implementation of existing government programmes.

The aims of *PAnKH* are very similar to the aims of the major existing government programmes such as SABLA[2] and RKSK, as is its community-based approach focused on engaging adolescent girls through group activities. However, in designing *PAnKH* we hoped to address some of the challenges faced by the implementation models that are currently being used, especially challenges related to the capacity, training and support of the young people expected to run the groups. First, *PAnKH* utilises para-professional mentors who are

---

[2] Rajiv Gandhi Scheme for Empowerment of Adolescent Girls

paid a competitive salary for their work with the idea that this professionalisation would increase commitment to the challenging role. Second, the mentors were also chosen to be slightly older than peer educators in government programmes (19 to 25 years rather than, e.g., 15 to 18 in RKSK) due to the demanding nature of the role. Finally, the programme provided intensive training, 28 days in total, to mentors and more frequent supervision and support.

Engaging with girls' wider community is an important feature of *PAnKH* and the research design we adopted allows us to assess the additional effects of this engagement over and above the activities focused on adolescent girls. This dimension of *PAnKH* is related to the approach of BBBP but following a more structured approach. Engagement primarily took the form of community 'Call for Action' events run by the adolescent girls and their mentors every two to three weeks. In these events girls and mentors presented the topics they had been covering in the groups to the community, facilitated a wider discussion and sought support and action from them on these issues. This gave these events a natural structure which closely followed the curriculum of the girls' groups.

In order to evaluate the effectiveness of *PAnKH*, we designed a cluster randomised controlled trial (RCT) in a way which allows us to not only learn about the overall impact of the *PAnKH* programme but also to assess whether engagement with the broader community makes the programme more effective. To this end we conducted a three arm trial in which a total of 90 clusters were randomly allocated to one of three arms: one in which only the activities directly targeting the girls (group education and sports sessions) were implemented, one in which these were complemented by community engagement activities and one in which none of the *PAnKH* programme components were implemented (the control group). We collect data in all three arms before and after the implementation of the programme to evaluate and compare impacts of the "girl only" and "integrated" models of the *PAnKH* programme on key outcomes including marriage, education, life-skills, mental health and gender attitudes.

The impact evaluation allows us to answer the following questions:

- What is the impact of the PAnKH integrated community-based programme on education, marriage, life-skills, mental health and gender attitudes of adolescent girls, compared to girls not receiving the programme?
- Are there benefits to targeting the girls' parents, other members of the community and local stakeholders in addition to the adolescent girls themselves?
- What impact does the programme have on hypothesised mechanisms for programme effects including knowledge, attitudes and life-skills?

These are the questions we focus on in this report.

Given our aim to provide a blueprint for scalable and cost-effective programmes which could be adopted within existing government policy frameworks, we offer in-depth discussion of the policy context and lessons learned about implementation, alongside detailed cost-effectiveness estimates in the companion *Programme Implementation and Cost-Effectiveness Analysis* report, which addresses the following key questions:

- How do the aims and design of *PAnKH* fit into the current policy landscape in India?
- How effective was the *PAnKH* programme at reaching the target group?
- How do the costs and benefits of the girl only approach compare to those of the integrated approach and to other programmes targeting similar outcomes?
- How could the *PAnKH* model be implemented at larger scale?

2   ROLE OF THE FUNDING SOURCE

The study was funded by MacArthur Foundation, International Inspiration and Children's Investment Fund Foundation The funders had no role in study design, data collection, analysis or interpretation of results.

3   THE PAnKH PROGRAMME

The primary aims of the *PAnKH* Programme were to delay age at marriage, increase school retention, foster the development of life skills, improve mental health and promote positive gender attitudes of adolescent girls in India. Moreover, additional goals included the creation of a supportive family and community environment that contributed to girl's development.

3.1   PROGRAMME COMPONENTS

The key components of the programme included:

Group Education Activities with adolescent girls age 12-19 (GEA): These consisted of a total of around 30 sessions and 48 activities, each activity lasting about 45-60 minutes. These were facilitated by "mentors" selected from amongst young women living in the targeted communities and trained to follow a curriculum developed by PRADAN and ICRW. The curriculum was divided into three segments: basic, intermediate and advanced comprising 11, 12 and 5-8 sessions, respectively (see Appendix D for a sample session). Mentors applied a variety of strategies in leading the sessions that involved interactive games, role-play and group discussions. Advanced curriculum was taught in separate sessions for older and younger, as well as, married and unmarried girls so that discussion of more sensitive topics related to the female body, sexual and reproductive health and family planning could be tailored appropriately. Moreover, the programme designed a Pankh Diary. This tool was created to summarize the positive massages of the programme and allow the girls to review them beyond the programme sessions.

Sports Activities with adolescent girls age 12-19: Building on ICRW experience of implementing *Parivartan* Girl Sport programme[3], as well as evidence from developed

---

[3] *Parivartan* Girl Sports programme is an intervention targeting adolescent girls age 12-16 years with the aim of improving their self-esteem, self-efficacy and aspirations for education:
http://strive.lshtm.ac.uk/resources/parivartan-girls-programme-tools

countries that sports activities can improve academic performance and school attendance among adolescents (Stevenson, 2010; Cappelen et al, 2018) a core component of work with the girls included weekly sports sessions of 60 minutes. These often took place on the same day as the education sessions – girls played a traditional sport called *Kabaddi*, while mentors initiated discussion of key issues covered in the education sessions.

Kabaddi tournament: Kabaddi sport tournaments were held during the project period providing an opportunity for girls from different villages to come together and interact with each other as well as other community members.

Call for Action Events: these consisted of events for the whole community at the end of each module of the GEA sessions. Girl invited their families and other members of the community and presented the material they had covered during the education sessions, discussing topics related to education, school drop-out, early marriage, social norms and gender oriented violence with members of the community.

In addition to this set of activities, the original design also included (i) activities with mothers and fathers of adolescent girls, and (ii) education activities with men and boys. There were some serious obstacles encountered in initial attempts to implement these[4] so it was decided to concentrate on the sub-set of core elements described above and focus on executing those well.

3.2 Programme Theory of Change

We propose a theory of change which outlines the main channels through which the *PAnKH* programme may affect marriage, education, life skills, mental health and gender attitudes. It is guided by our overarching hypothesis that *adolescent girls will be able to make critical choices relating to marriage, education, health and livelihood only if they have the right information, understanding and skills to navigate their contexts, are free from fear and violence, are valued by society, and live in a supportive environment both within and outside home.* We identify five specific

PAnKH which means "wings" in English is symbolic of the aspiration to be able to chart the course of journey of one's own life. Wings are related to the cognitive faculty, imagination, thought, freedom and victory.

We perceive *"PAnKH"* as an expression of empowerment, where adolescent girls are able to explore their maximum capabilities and fulfill all their aspirations by making independent choices at critical junctures such as taking decisions on their health, wellbeing, education, livelihood, choice of partner and reproduction.

 We envisage that our project will be instrumental in creating a supportive (and safe) environment in communities so that the girls are valued by the society and a girl's act of challenging and breaking free from the inequitable social norms does not jeopardize her basic right to freedom from fear, gender discrimination and gender based violence

---

[4] In particular, attendance in the few men and boys, as well as family sessions that did run was very low.

potential mechanisms for programme effects. These include (i) improvement in knowledge; (ii) changes in attitudes; (iii) increased social support; (iv) enhanced life skills and mental health; and (v) creation of enabling environments within the household and community.

We now discuss the role and importance of each mechanism in turn:

Improved Knowledge: Lack of knowledge is likely to be a key constraint in adolescent decision-making, as evidenced, for example, by widespread health campaigns targeting adolescents about the danger of risky behaviours. We expect lack of knowledge about sexual and reproductive health and relationships to be a particularly salient constraint in the study context, where these issues are considered taboo and inappropriate to discuss. Providing adolescents with accurate information related to education (e.g. benefits of educations, rights in school), marriage (e.g. legal age of marriage, relationships in a marriage) and sexual and reproductive health (e.g. information about menstruation and contraception) through the *PAnKH* programme may shift girls' perceptions of costs and benefits of different decisions related to these areas.

Changes in Attitudes: By introducing new ideas and concepts around gender, education, marriage and sexual and reproductive health, as well as encouraging girls to critically engage with entrenched norms, the programme altering girls' attitudes to key life decisions. Having a clearer conception of some of the prevailing norms and attitudes in negotiation and decision making within the household may help girls to overcome the barriers these attitudes and norms form. For example, by promoting a pro-active attitude to involvement in decision-making within marriage (along with strategies to do so) the programme may increase the amount of say girls have about use of contraception (a taboo issue which men usually have final say on).

Increased Social Support: Peers and the wider social setting are a particularly important influence in decision-making and wellbeing in adolescence (Knoll et al, 2015). By bringing together adolescent girls in a safe environment, the programme will encourage the formation of new networks and friendships, which will increase the social support and information available to the girls. These networks may also change girls' perceptions of 'norms' in the community and influence attitudes. Programme influence on the norms and attitudes of the other members of the community (mothers, men & boys, wider community) may further enhance the support available to girls in their homes and within the community.

Enhanced Life Skills and Mental Health: Adolescence is increasingly recognised as a critical stage for the formation of life skills (such as self-efficacy, self-esteem, peer-relations and socio-emotional skills) and mental health which are crucial for lifelong wellbeing (Kia-Keating et al. 2011), educational attainment, skills and socio-economic outcomes (Heckman et al. 2006). The *PAnKH* curriculum targets these critical domains through sports, activities and discussion within the education sessions, along with the promotion of role-models, which encourage girls to feel positive about themselves and support the idea that girls like them have the abilities to succeed in various challenging situations and to affect change. Previous studies have found that interventions with similar components – sports (Ekeland et al. 2005; Dishman et al. 2004) and group based reflective discussions and other games and activities

(see Morton and Montgomery (2013) for a review) – had impacts on self-efficacy, self esteem and other measures of life skills and mental health.

Creation of Enabling Environments within the Household and Community: The outcomes of adolescent girls in our study environment are affected by many factors outside of their direct control. Attitudes, norms and practices in the family and wider community are key. These are targeted by the programme through engagement not just with the girls but with members of their households and the wider community. By working with these groups to consider and question traditional gender norms within the community and in the natal and marital home the programme may make the environment in which the girls live more receptive to their needs, open to their input and safer for them, enabling and supporting them in making the best life choices for them.

Detailed description on implementation of the PAnKH programme can be found in the companion *Programme Implementation and Cost Effectiveness Analysis* report.

4   RESEARCH DESIGN

In order to test the effectiveness of *PAnKH*, we designed a randomised controlled trial (RCT) of the programme. The aim of the trail is to evaluate the overall impact of the *PAnKH* programme, as well as to disentangle the effects of programme components that target only the girls from those that also engages with the broader community. In this way we can assess not only the impact and cost-effectiveness of the full integrated model of the programme, but also of its key individual components and, potentially, learn more about mechanisms for programme impacts.

To this end we conducted a three arm trial in which a total of 90 clusters were randomly allocated to one of three arms: one in which the full integrated model of the *PAnKH* programme was implemented, one in which only the activities directly targeting the girls (group education and sports sessions) were implemented and one in which none of the *PAnKH* programme components were implemented (the control group). We collect data in all three arms before and after the implementation of the programme to evaluate and compare impacts of the "girl only" and "integrated" models of the *PAnKH* programme on the main outcomes identified in the Theory of Change. In this section we describe the evaluation design, sampling procedures, data collection and outcome measures.

Ethical clearance for the project was granted by the Sigma IRB Committee (New Delhi), ICRW IRB (Washington DC), and Oxford University IRB (Oxford, UK).

4.1   EVALUATION DESIGN

Evaluation Problem: The key evaluation problem is that we would like to observe how the same individual would fare with and without the programme; for instance, we would like to compare the age at marriage of a girl in a state of the world where she has participated in the *PAnKH* programme to that where she has not. Since it is not possible to observe these two

outcomes for the same girl we need to use an approach that allows us to observe the outcome for two groups of girls where:

1. Only one group participated in the programme
2. Without the programme, the outcomes *would* have been identical.

As long as these conditions hold, any differences in outcomes found after programme implementation can be attributed to the programme. Because participating in the programme is an individual choice, it is likely that girls who choose to participate are different from those who choose not to participate, so that even without the programme their outcomes would not have been identical. A simple comparison of participants and non-participants would, therefore, not allow us to disentangle programme effects from these pre-existing differences.

The first-best way to address these pre-existing differences is to allocate programme participation *randomly* among a group of eligible individuals. Conditional of successful randomization and a large enough sample, allocating the programme in this way will mean that girls who are assigned to the *treatment* group (group that receives the programme) are, on average, not different, before the start of the programme, from the girls who are in the *control* group (group that does not receive the programme), meeting conditions (1) and (2) above. In this case any differences that are found after implementation of the programme can be confidently attributed to programme impacts. This is the approach that we adopt to evaluate the impact of the *PAnKH* programme.

It is important to note that this is a first best design for informing on the impact of *being in a village where the programme was implemented* (Intention to Treat), rather than the impact of *directly participating in the programme*. These are not the same unless all of those eligible for the programme participate in the programme. Using an *Intention to Treat* framework is optimal in this study for two reasons. First, if we were to focus on those directly participating in the programme we risk introducing systematic differences between the treatment and comparison groups, to the extent that girls/households who chose to participate were already different from those who chose not to. This would mean that although the programme was randomised, we would not be able to confidently attribute any differences found between the treatment and control groups at the end of the programme to the effects of the programme as they could (in part) be driven by pre-existing differences. In addition, the *unconditional* effect of a household being offered the programme is arguably more useful from a policy perspective at it is our best estimate on the effect on the 'average' adolescent girl of scaling up the programme.

Randomisation strategy: Our unit of randomisation is the village, since the programme operates at the community level. We have a total sample of 90 villages and randomly allocated 30 of the 90 villages to each of three treatment arms:

1. "Integrated", in which the full set of *PAnKH* programme components are implemented (education and sports sessions for girls + call for action events and community mobilization and social campaigning)

2. "Girl only" in which only the activities with the girls (education and sport sessions) are implemented.
3. "Control" in which none of the *PAnKH* programme components are implemented

Evaluation approach: In order to evaluate programme impacts data was collected at the start (baseline) and end (endline) of the project. Baseline data were collected before the start of programme implementation between October and December, 2015. Endline data were collected between December 2017 and March 2018. We use these data to estimate the impact of eligibility for the *PAnKH* programme on the targeted outcomes using regression analysis and controlling for a core set of baseline characteristics to increase the precision of the impact estimates. The empirical analysis strategy is set out in more detail in Section 6.1.1. This can also be found in the pre-analysis plan (Appendix C) which was published online prior to completion of endline data collection.[5]

Due to resource constraints the evaluation focuses on measuring impacts on the primary target beneficiaries of the programme – the adolescent girls. To this end we collect data on the girls, their caregivers and households. We are not, therefore, able to measures direct impacts on men/boys or wider set of community members reached by the programme. This would have been especially relevant in the "integrated" arm where the wider community was directly targeted.

4.2   EVALUATION IMPLEMENTATION

4.2.1   STUDY SAMPLE

Selection of districts and blocks: The trial took place in three blocks of Dholpur district – Bari, Basari and Dholpur. The sites selected were those where PRADAN, the programme implementer, has operated for over ten years and has built up the well-established self-help group infrastructure and trust of the communities which were essential for successful implementation of the PAnKH pogramme[6].

Study Sample Selection: The PAnKH programme targeted girls age 12-19. This means that all girls in this age-range living in programme villages were eligible to participate. Within this age range we identified 3 groups of girls who might be affected differently by the programme with the aim of sampling in a way which would allow us to study impacts separately for each of these groups. The groups include (1) younger adolescent girls (12-14 years); (2) older, unmarried adolescent girls (15-17 years); and (3) married adolescent girls (12-19 years). The baseline sample, drawn from a list of all 12-19 year old girls living in the study villages at the start of the project (compiled through a comprehensive listing exercise conducted as part of this study), included a total of 7,577 girls. It was made up of 3,096 12-14 year old unmarried

[5] https://www.ifs.org.uk/publications/10347
Outcome variables listed in online registrations: http://www.isrctn.com/ISRCTN77449378 and https://www.socialscienceregistry.org/trials/2656
[6] More in-depth discussion of the rationale for site selection can be found in the companion *Programme Implementation and Cost Effectiveness Analysis* report.

girls, 2,827 15-17 year old unmarried girls and 1,654 15-19 year old married girls (there were no girls younger than 15 reported as married), roughly equally split across the three study arms. Analysis of sample balance showed very few differences in core characteristics of the girls across the three study arms, suggesting that the randomisation was successful and, before the start of the programme, girls who were assigned to the treatment arms were on average not different from each other along observable characteristics than the girls assigned to the control group.

Further details of study sample selection and balance at baseline can be found in the baseline report (Achyut et al. 2016).

Study Sample Follow-up: The key difference between the baseline and follow-up samples is the exclusion of the 1,654 girls who were married at baseline. Only a small minority (7%) of them participated in the programme which meant that we would not have sufficient statistical power to detect programme impacts on this sub-group. We, therefore, decided to not follow them up as part of the endline data collection. Instead we conducted qualitative work with a sub-sample to investigate key barriers to programme participation for this extremely hard to reach group. The results of this qualitative investigation are reported in the companion *Programme Implementation and Cost Effectiveness Analysis* report.

Therefore, at endline we aimed to re-interview 5,921 girls who were age 12-17 and unmarried at baseline. We were able to collect complete endline data for 5,150 girls (87% of the sample) and partial data through parental report for 96.3% of the sample. The full trial flow diagram is presented in Figure 1. An attrition rate of 13% compares well to other studies that target adolescent girls in developing countries in a similar way in which attrition tends to be closer to 20% (e.g. Adoho et al, 2014; Buchmann et al, 2017; Bandiera et al, 2017).

Table 1 shows that attrition rates and reasons for attrition were similar across the 3 arms. The three most common reasons for attrition included migration outside Dholpur district, girl not being available for the interview repeatedly (fieldwork protocol stipulated 3 revisits), refusal to participate in the survey by girl or family. We test for differential attrition by treatment status and find no evidence to suggest that this is an issue (Appendix Table B2). In line with this, analysis of balance of the endline sample on baseline characteristics shows that no imbalances emerge as the result of attrition; in other words while the imbalances identified during baseline persist in the endline sample, there are no new imbalances (Appendix Table B1). As specified in the pre-analysis plan, we address baseline imbalances in our impact analysis by adding controls for the characteristics found to be imbalanced in the baseline.

Table 1: Sample size and attrition by treatment status & reason for attrition

	Girl Only	Integrated	Control	Total
Sample at baseline (unmarried girls)	1919	1966	2036	5921
Sample at follow-up	1665	1704	1781	5150
Attrition rate	**13.2%**	**13.3%**	**12.5%**	**13.0%**
Reasons for Attrition				
Girl moved outside Dhaulpur or to area difficult to reach	141	128	134	403
Girl died	4	6	9	19
Household not located	9	11	15	35
Refusal	43	37	24	104
Girl unavailable after 3 visits	57	80	72	209
Other	0	0	1	1

Figure 1: Trial Flow Diagram Profile
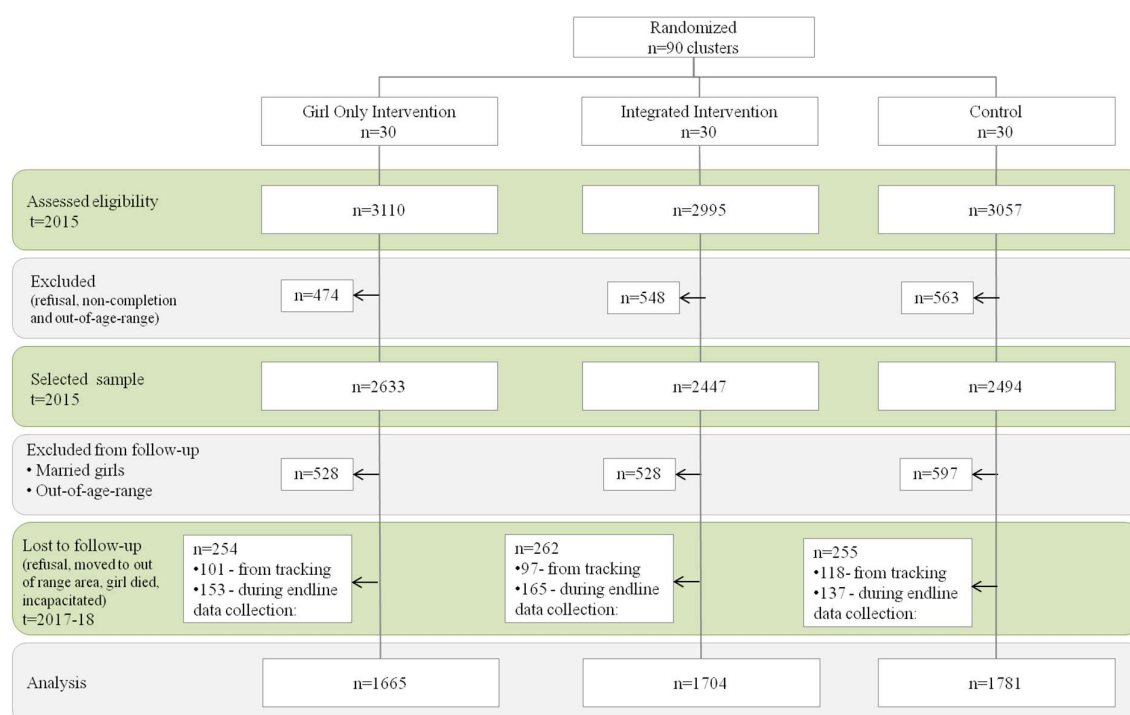
4.2.2 DATA COLLECTION

Data collection took place between October and December 2015 for baseline and December 2017 and March 2018 for endline. Endline was preceded by a tracking exercise between October and December 2017, during which a big effort was made to locate all of the non-married girls from the baseline sample, including those who had moved within the district.

Questionnaires administered: The questionnaires administered during baseline and endline data-collection include:

(1) <u>Girl questionnaire</u> was administered to all girls in the sample eligible for the *PAnKH* programme. It contained questions on the following broad themes: education, employment, savings, time-use, peer-network, self-esteem, self-efficacy, resilience, mental health, attitudes towards gender norms, marriage, sexual and reproductive health and family planning (knowledge and practices), safety and violence, communication and well-being, access and knowledge of *PAnKH*.

(2) <u>Carer questionnaire</u> was administered to the main carer of each girl in the baseline sample. It contained questions on the following broad themes: education, employment, decision making, attitudes to girls' education and marriage, menstruation and sexual and reproductive health knowledge, child-related expenditures within the household, social support, attitudes towards gender norms.

(3) <u>Household questionnaire</u> was administered to a knowledgeable adult in the household and contained questions on the following broad themes: a household roster, dwelling characteristics, assets, ration cards, caste and religion of the household members.

(4) <u>Mentor questionnaire</u> was administered to 120 women who worked as *PAnKH* mentors, facilitating the education and sports sessions. This questionnaire was only administered at endline and included questions on the following themes: education, employment, peer-network, self-esteem, self-efficacy, resilience, mental health, attitudes towards gender norms, marriage, sexual and reproductive health, family planning, safety and violence, and knowledge of *PAnKH*.

Since some girls had moved from their baseline households by the time of the endline, the following guidelines were followed in administration of the questionnaires:

- If at endline the target girl was still living in the baseline household, the interviewer administered the girl, carer and roster section of the household questionnaires. For the cases where the girl had no carer, only the girl questionnaire and roster of the household questionnaire were administered.
- If by endline the target girl had moved from her baseline household, the girl questionnaire and a full household questionnaire were administered in the new household. If the caregiver was still alive and available in the baseline household, the carer questionnaire was administered in the baseline household. In the cases where the girl had moved to a hostel or rental place, the household questionnaire was not administered.

Data collection procedures: We conducted a competitive bidding process to select the survey company for endline data collection. We received technical and financial proposals from six companies and shortlisted three on the basis of independent ratings by the ICRW and IFS teams. The three shortlisted companies presented their proposals to the ICRW research team. On this basis, Change Alliance won the bid for the endline data collection contract.

Baseline data collection procedures are detailed in the baseline report.

The endline data-collection team consisted of 2 fieldwork leaders, and 8 teams with 5 field staff in each (4 female interviewers and 1 male supervisor). For ethical reasons and to ensure that girls felt comfortable and able to give honest answers to sensitive questions, all investigators were female. The fieldwork leaders were in charge of assigning a set of interviews to each supervisor, monitoring the entire data collection process, keeping records of the status of interviews and approving quality and completion of interviews. The supervisors allocated households to each of the investigators (interviewers) and were responsible for monitoring interview quality following an agreed process. The team also consisted of 4 female quality monitors from the survey company (Change Alliance), 2 female consultants and a team of 4 researchers recruited by ICRW to provide further quality monitoring. IFS provided support by monitoring data collection progress and identifying cases with missing data that needed additional attention.

Before starting data collection, all fieldwork staff received 11 days of training which included: one day focusing on the aims and objectives of the study, general interviewing protocols, and ethical research guidelines including consent and assent processes; three days familiarising themselves with the household and girl questionnaires and the use of tablets for data collection[7]; four days on the carer questionnaire; two days for field practice, debriefing and discussion.

Once in the field each interviewer was assigned a set of sample households by her supervisor. Each household visit would start with an introduction and asking to speak to the household head. The interviewer then discussed the purpose of her visit with the household head and explained the household consent form, responding to any follow-up questions. The interviewer only proceeded with the household interview if she was able to secure written consent to participation in the study from the household head. Upon completion of the household survey, the interviewer approached the selected eligible girls and their female caregiver for the individual surveys[8].

Quality monitoring: To ensure collection of high quality data we implemented three categories of monitoring mechanisms which are summarised in Figure 2 below. These

---

[7] All data were collected using Computer-assisted personal interviewing (CAPI).
[8] Depending on the age of the selected respondents, the interviewer approached a parent (or respondent, in case of adult) for their consent and then the respondent (girl) for assent.

include a) guidance and revision ex-ante/ex-post interview; b) supervision during interviews; and c) field re-visits and back-checks.

Overall the quality monitoring process ensured that the correct respondent/household was interviewed, following the guidelines above; that all ethical procedures were followed to an acceptable standard; that data collected was of good quality; and that interview refusals or failure to find sample households were minimised. Key procedures in place included quality monitoring during interviews by monitors sitting in on interviews, partial re-interview of a random 10% of completed interviews by supervisors, periodic review of all uploaded collected data which provided real-time information on the quality of the data being collected used to collate reports for the field team on what additional checks were necessary for already collected data and what to pay more attention to in ongoing data collection. Finally, there was multi-stage follow-up on any incomplete interviews including up to three re-visits to the household to try to complete the interview.

Figure 2. Data collection and quality control process
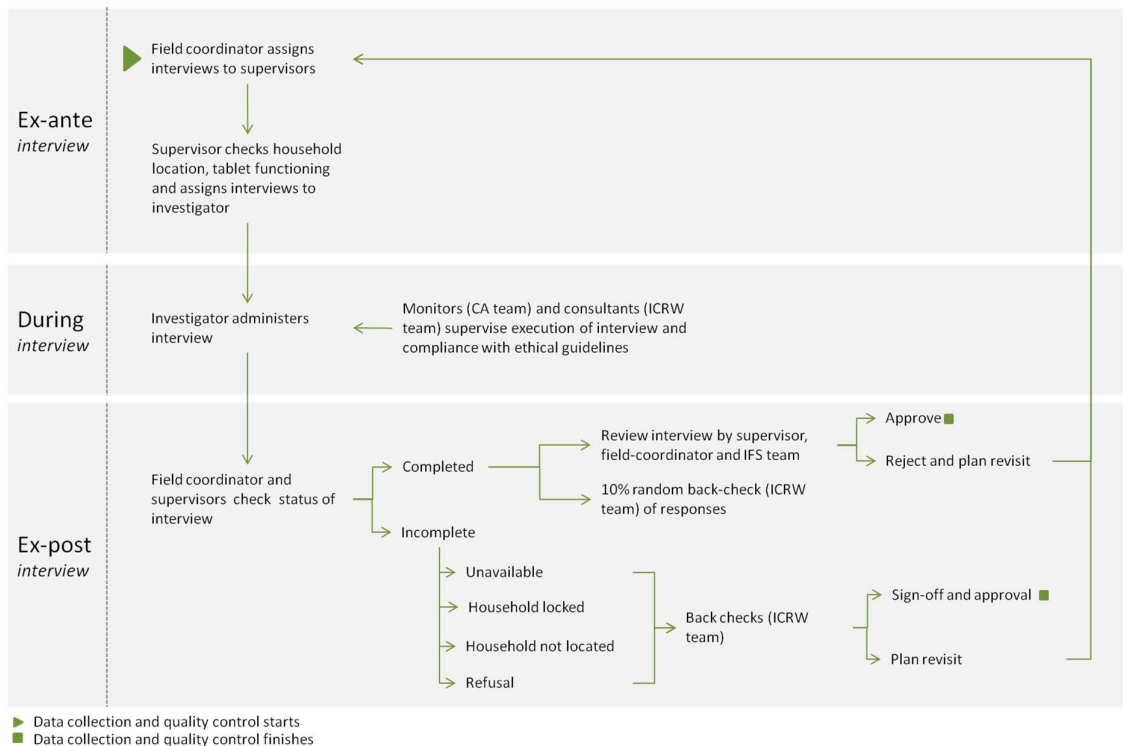
4.2.3 MEASUREMENT

We selected primary and secondary outcomes on the basis of (i) PAnKH's aims; and (ii) power calculations of 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 is possible to detect realistically sized treatment effects given our sample size and study design.

Many of our outcomes are measured by asking respondents to indicate their level of agreement with sets of statements relating to a given construct or practice. We were careful to where possible (a) select from existing scales; and (b) select measures which have been shown to have adequate psychometric properties in settings as similar to ours as possible. The final selection of instruments was made in consultation with Dr Marc Bornstein, an expert in cross-cultural research into child and adolescent development.

Our primary outcomes include indicators of whether the target girl is married, whether she is attending school at the time of the endline, age-specific marriage and school attendance rates, as well as measures of mental health and life skills. Mental health is measured using three separate scales:

- Patient Health Questionnaire (PHQ) to capture depression, validated for use in adults and adolescent girls in India.
- Generalised Anxiety Disorder (GAD-7) scale, which has also been adapted for and validated on a population of adolescent girls in India.
- Rumination Responses Scale developed to measure rumination (repetitive and passive thinking in response to depressive moods). While the scale had previously been used in many contexts and with adolescents, it had not previously been administered to adolescent girls in rural Rajasthan so had to be translated adapted and piloted extensively for this study.

We also measure a range of life skills including self-efficacy, self-esteem, peer relations, resilience, vigilance in decision making (tendency to clarify objectives, search for information and carefully evaluate alternatives) and buck-passing in decision making (tendency to avoid making decisions/leave decisions to others). While the self-efficacy, self-esteem, peer relations and resilience measures had been used in the Indian context with adolescents in previous studies, the decision-making scales had not so needed to be translated, piloted and adapted for this study.

Our final primary outcome is gender attitudes which we measured using a scale partially based on the Gender Equitable Men Scale (GEMS) which has been used in India in several studies. While the full scale has 29 items, exploratory factor analysis of the baseline data showed that 14 of these load on a single unidimensional measure of patriarchal gender attitudes, which is of primary interest in this study. Our outcome measure of gender attitudes is, therefore, constructed from a sub-scale of the GEMS containing these 14 items administered to the girls as well as their caregivers.

Additionally, we define a set of secondary outcomes that, according to our theory of change, reflect potential mechanisms through which *PAnKH* may have an impact. The first of these is attitudes to school, measured using a nine-item scale created by the research team asking the respondent to indicate degree of agreement with statements such as "I learn a lot of new things in school" and "I feel my studies have no meaning". The second is knowledge of sexual and reproductive health - measured using a puberty and menstruation knowledge scale

and a contraception and sexual health knowledge scale. Both are a set of multiple choice and true/false questions assessing knowledge that could be gained from experience (for example, normal duration of cycle and blood flow), as well as that which has to be learned (for example selecting true/false for statement relating to "menstruation cleans the body of dirty blood").

As discussed in the baseline report, we struggled to measure experience of violence. Although qualitative data suggest strongly that women and girls experience substantial levels of violence and harassment, almost none were reported in response to questions directly asking about such experiences. Follow-up focus-group discussions suggested that this was likely due to culture of "blaming" victims of violence for behaving in a way which "induce" violence against them. Following extensive piloting, at endline, we, therefore, replaced questions about own experience of violence with 6 different vignettes, each followed by 6 yes/no questions. The responses allowed us to construct measures of 6 domains: (1) victim blaming tendencies – the degree to which girl blame the victim for inducing violence against herself; (2) perpetrator blaming tendencies – the reverse of (1) i.e. the extent to which girls blame the perpetrator rather than the victim; (3) avoidance behaviour – tendency to deal with violence by avoiding situations which might incite violence (e.g. avoiding places where previously harassed); (4) retaliation behaviour; (5) reporting behaviour in response to violence.

The final two outcomes include girls' attitudes to sport and restrictions faced during menstruation. Girls were asked about their attitudes to sport using a 10 item scale with statements such as "I can play sports with the other girls in my village" and "I can compete in sports just as well as boys". The scale about restrictions faced during menstruation is a set of yes/no questions about being allowed to undertake different types of activities which girls are traditionally not allowed to do during menstruation. These include, for example, attending religious functions, touching family members and visiting relatives[9].

Outcome measures were constructed using appropriate scoring procedures. Scales with Likert type response options (e.g. Strongly Agree, Agree,...) were scored using exploratory factor analysis (iterated principal factor) to construct factor score(s) for each scale. Scales with binary response options (e.g. yes/no, true/not true) were scored using item response theory (IRT) two parameter model. We used a latent factor model to estimate the difficulty and discrimination parameters associated with each item, predicting the underlying score using an empirical Bayes methodology. More details on the scoring procedures can be found in the pre-analysis plan (Appendix C).

5 STUDY CONTEXT

---

[9] A detailed description of all of the outcome measures can be found in Tables 3 and 4 of the pre-analysis plan (see Appendix C).

Before discussing the main evaluation results we use our baseline and endline data[10], combined with the data from the 'Youth in India' representative study of youth in India[11] to contextualize the findings – describing trends that are most relevant for this study. This is particularly important for our study context since it has not been studied extensively and there are few data sources available.
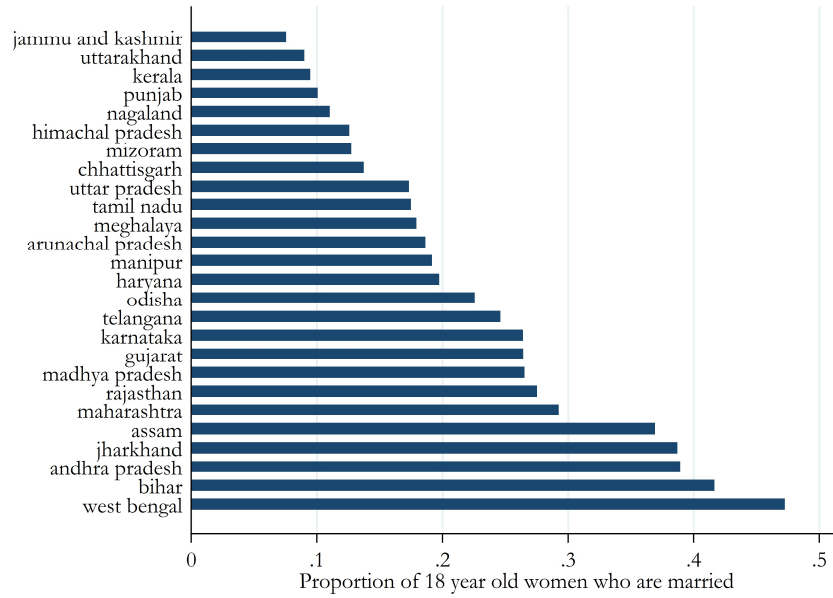
Marriage

Although there have been dramatic improvements in reported rates of early marriage (before age 18) in Rajasthan, it remains one of the worst performing states in terms of the proportion of 18 year-old women who are married. Based on the most recent National Family Health Survey (NFHS-IV) data, on average, the state of Rajasthan has the seventh highest rate of early marriage, with just over a quarter of 18 year old women reporting to be married in 2016/17. By contrast, this proportion is under 10% in the best performing states such as Kerala and Uttarakhand (Figure 3). However, there is reason for optimism as the rate of improvement over the last decade especially has been remarkable. Figure 4 shows that between the 2006 and 2016 rounds of NFHS proportion of 17 year old girls reporting to be married went down from almost half to just under 15%.

---

[10] We present baseline and endline descriptive statistics that correspond to the sample of girls that was follow-up and with completed data at endline. We do not have a representative sample and, therefore, the descriptive statistics presented are not representative of the population averages in the surveyed areas.

[11] The 'Youth in India' study collected data for 50,848 youth aged 15-29 years old in 6 states of India: Andhra Pradesh, Bihar, Jharkhand, Maharashtra, Rajasthan and Tamil Nadu. The study 'was designed to provide estimates for the state as a whole, as well as for urban and rural areas for each of the four categories of respondents, namely married and unmarried young men and women, separately' (IIPS and Population Council, 2010).

Figure 3: Propotion of 18 year old women who are currently married (NFHS-IV, 2016/17)

Figure 4: Proportion of women in rural areas of Rajasthan 'currently married' over 4 rounds of the NFHS
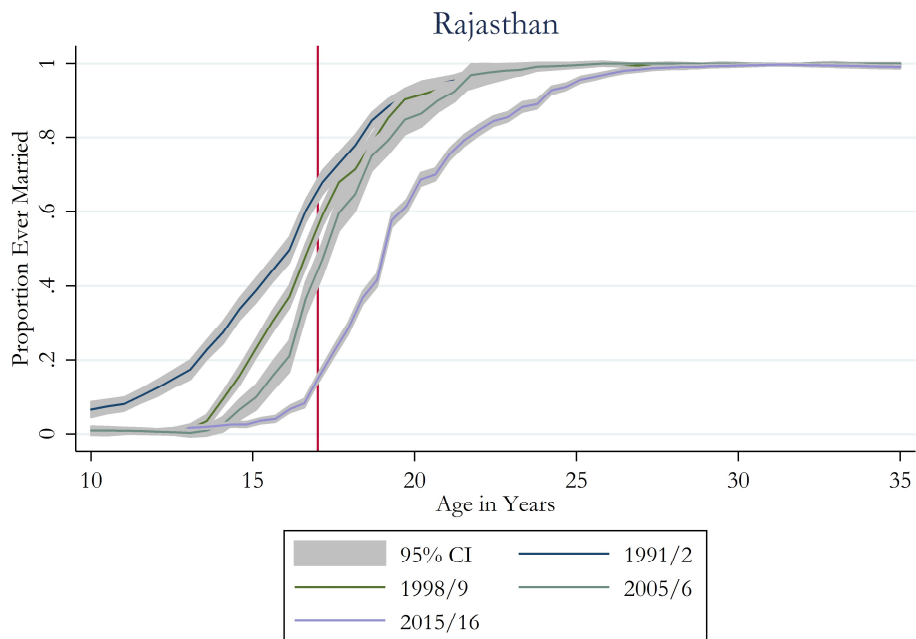
One of the main goals of PANKH was to delay marriage. The census data collected in 2015 in the sampled villages, showed that the proportion of girls reported as married was very low between ages 12 and 14, increasing from age 15 and reaching 13% of girls by age 17. By age 19, two thirds of the girls were reported as married.

Among the 12-17 year old unmarried girls in the study baseline sample, 12.6% were married by the time of the endline[12]. On average, girls got married to boys who were 3 years older than them and had 1.7 more years of education. According to parental report girls tended to marry boys from families that had the same wealth as them.

Further, as in baseline, the overwhelming norm was for arranged marriage: all of girls who got married between baseline and endline had an arranged marriage. The great majority of them (94%) had also said during the baseline survey that they would prefer an arranged marriage. The power to make marriage decisions was predominantly held by the father and girls had little say with respect to whom to marry and when to marry.

This is consistent also with the expectations of the girls who were still unmarried at the time of the endline: as in the baseline, nearly all of them expected the father or other relatives to have the maximum say in their marriage arrangements.

Education

In line with the marriage trends, Rajasthan is also one of the worst performing states with respect to girls' schooling. The 45% school attendance rate among 18 year old girls here contrasts sharply with Kerala, for example, where the attendance rate is double that at 90% (Figure 5). However, again, there has been a marked improvement, especially, over the last 10 years (Figure 6).

---

[12] This includes 1% who were engaged but not yet married.

Figure 5: Propotion of 18 year old women who are currently in school (NFHS-IV, 2016/17)

Proportion of 18 year old women in school
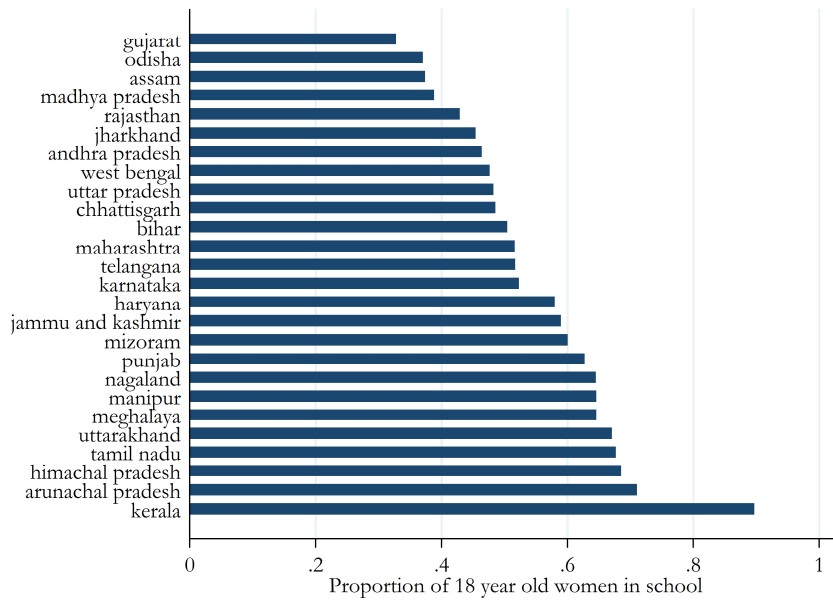
Figure 6: Proportion of women in rural areas of Rajasthan 'currently in school' over 4 rounds of the NFHS
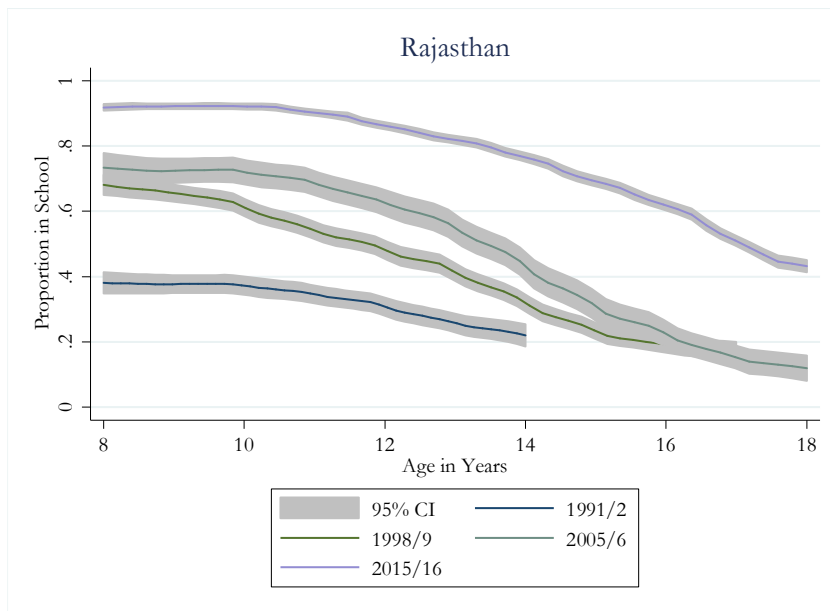
In our study sample, at baseline 76% of girls who were age 12-17 and unmarried were attending school. By endline, 31% of these girls had completed or stopped attending school. Figure 7 shows that at all ages reported school participation among married girls was much lower than among non-married girls. For example, while 61% of unmarried girls report attending school, this is the case for only 10% of married girls. Finally, the data suggest that the chances that girls return to school once they stop attending are very low – among girls who were not attending school at baseline, only 1% were attending school at endline.

Figure 7: School attendance by age and marital status at endline
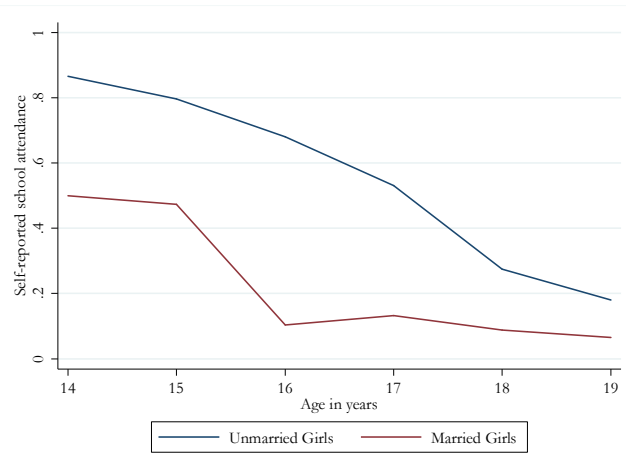
More than 90% of the girls who dropped-out of school between baseline and endline reached a lower level than the one they aspired to and expected to obtain when asked at baseline. For instance, nearly a quarter wished to achieve tertiary education (professional diploma, graduate and postgraduate studies) but did not even complete school. Why did this happen? Across married and unmarried girls, two fifths cited at least one of the following - domestic responsibilities, need to work and families being unable to afford school. Contrary to a widespread hypothesis that in contexts where early marriage is prevalent girls are taken out of school for marriage, 62% (or nearly two out of three) cited reasons other than marriage to explain why they left school.

Figure 8. Reasons for dropping out of school
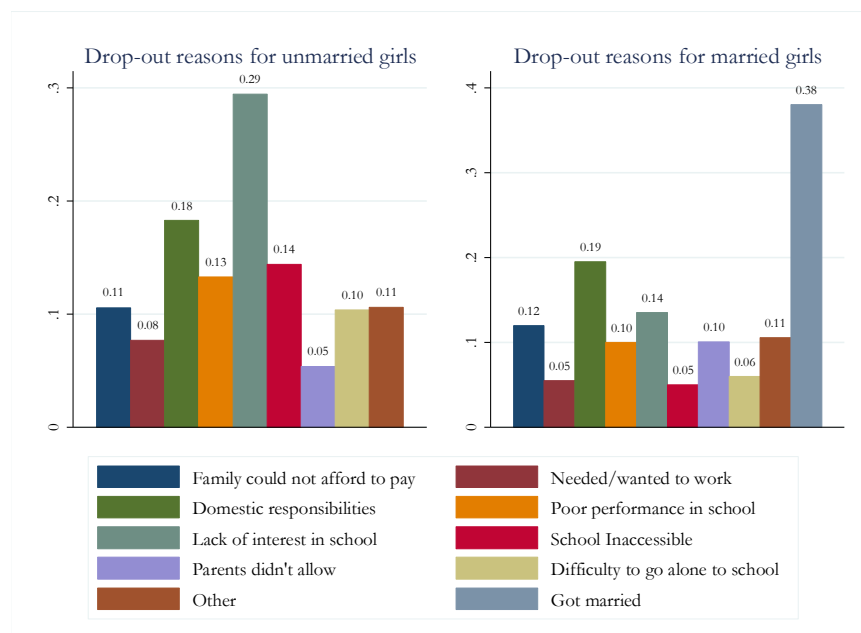
Mental Health

There is growing evidence that adolescents and new mothers experience heightened vulnerability to developing mental health disorders with adverse long-term consequences for themselves and their children. At the same time mental health is generally recognised as a neglected area when it comes to health service provision in developing countries (Stein et al, 2014; Kieling et al, 2011). Within the PAnKH programme mental heath is a key channel for impacting girls well-being and life outcomes.

Generally, data containing measures of adolescent mental health are rare in developing county contexts. Here we utilize data collected as part of this study as well as the 'Youth in India' data to look at some basic correlations between mental health and key factors such as age, marital status and socio-economic status.

The 'Youth in India' data includes the General Health Questionnaire (GHQ-12) to measure mental health[13]. The data suggest that mental health decreases with age: for example, girls aged 15 to 17 have better mental health compared to older girls aged 18-24. We also observe that women with a higher level of education have better mental health (Figure 9).

---

[13] To measure mental health we applied Item Response Theory to responses to the GHQ-12 scale and predicted a mental health score.

Figure 9: Education and mental health

Mental health by level of education

Notes: higher scores mean better mental health. Source: Youth Data 2006-2007
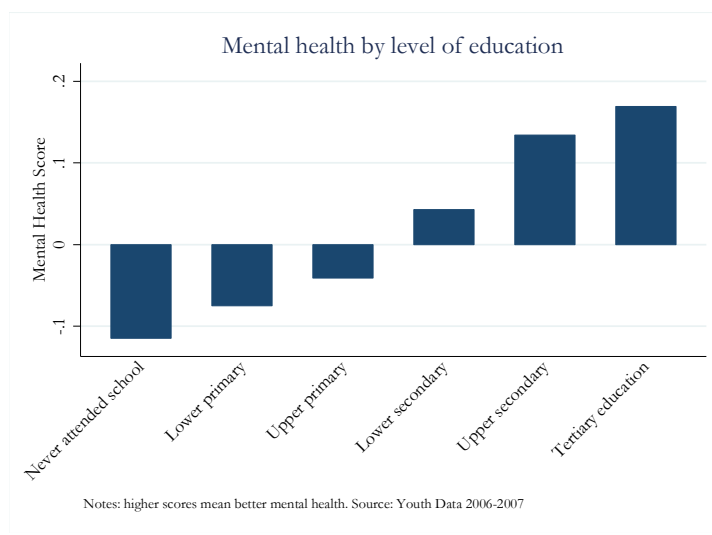
It is hypothesized that early marriage (marriage before 18) can influence the development of mental health disorders as it leads to an abrupt end of adolescence, early pregnancies, potential disruption of education, and a reduction of autonomy and mobility (UNICEF, 2001). This appears to be in line with correlations in the 'Youth in India' data which show that women married at 19-25 years of age have better mental health relative to women married at a younger age.

The endline of our study includes measures of three domains of girls' mental health including depression, anxiety and rumination. The instruments used to measure these are described in Section 4.2.3. In line with trends in the 'Youth in India' data, higher education and wealth are positively correlated with girls' mental health.

Gender attitudes

Attitudes to the role of women and men in the family and the community are a major focus of the *PAnKH* programme. Our measure of gender norms is described in Section 4.2.3. At both baseline and endline we find that the majority of girls expressed agreement with statements supporting traditional gender roles. For example, in both rounds around 2 out of 3 girls either agreed or strongly agreed with the statement "A man should have the final say in all family matters". Within our study sample, older unmarried girls appear to hold less patriarchal attitudes than their younger counterparts, as do girls from dominant caste households and higher wealth households.

Further, both in baseline and endline we find that gender attitudes of the girls are closely correlated with those held by their primary carers. However, we observe that the difference between carers' and girls' gender norms is bigger at endline than baseline, with girls holding less patriarchal norms compared to their caregivers (Figure 10). This change is observed in the control group so is not driven by the intervention; rather it may reflect an age effect.

27

Figure 10: Girl and Carer Gender Patriarchal Attitudes at Baseline and Endline
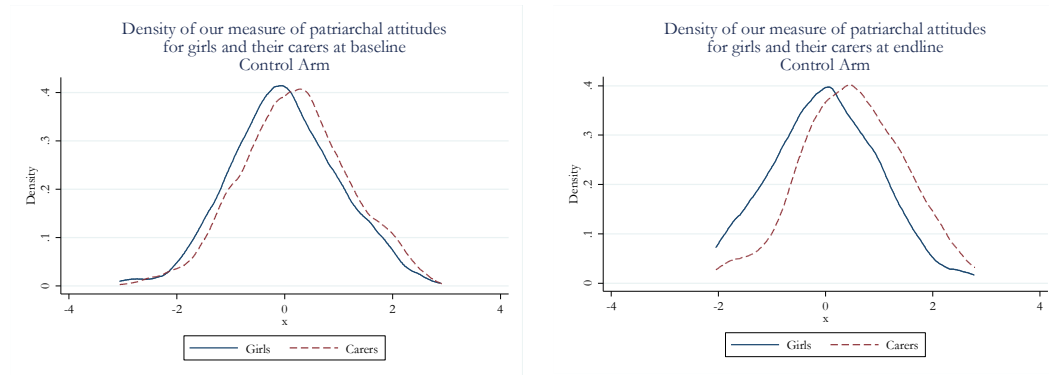
Menstruation, sexual and reproductive health

Sexual and reproductive health is a major issue for adolescents and young people in India, especially girls. Prevalence of misconceptions and myths, restrictive traditional practices, shame and stigma surrounding menstruation, combined with restricted access to suitable menstrual hygiene products and WASH facilities mean that adolescent girls are ill informed about menstruation, have few sources of information to draw on and perceive menstruation as a taboo subject. Existing studies on India and Rajasthan suggest that these factors are associated with unsafe menstrual hygiene practices with heightened risk of reproductive tract infections, as well as school absenteeism (Khanna et al, 2005; van Eijk et al, 2016)13. Further, prevalence of early marriage results in early onset of sexual activity, extremely low rates of contraception use, early childbearing and heightened vulnerability to maternal mortality and sexual violence (Jejeebhoy and Santhya, 2011). One of the aims of the *PAnKH* programme is, therefore, to improve girls' knowledge about sexual and reproductive health, as well as exploring strategies for having more say in family planning.

Menstruation

Our data show that this is a context where girls have little knowledge of the female body or the purpose of menstruation. For example, consistently, in both baseline and endline only a tiny minority (3-5%) of girls correctly disagreed with the statement that "Menstruation cleans the body of dirty blood" and only 1 in 4 14-19 year-olds at endline correctly disagreed with the statement that "Urine and menstrual blood leave the body through the same path". It is also clear that there is very little discussion of issues relating to menstruation and other changes that happen during puberty; according to the endline data only just over a third of the girls had ever talked to their parents about these issues.

There are also widespread culturally motivated restrictions placed on menstruating girls. The vast majority of girls in both survey rounds were not allowed to touch stored food or attend religious activities during menstruation; roughly 40% were not even allowed to cook.

Sexual and reproductive health

The level of knowledge about reproduction and family planning methods at baseline and endline was also low. Only about half of the girls age 15+ could name at least one contraceptive method without prompting; this proportion is higher among girls who are older, and those from households in the top wealth quartile. The main contraceptive methods named by the girls were female sterilization and contraception pills. Moreover, girls also showed knowledge gaps regarding the links between the menstruation cycle and the likelihood of pregnancy; only a quarter correctly answered that a woman is more likely to get pregnant from day 8 to 19 of the menstrual cycle.

Life skills

At baseline and endline we measured the girls' self-efficacy, self-esteem and peer relations. Both rounds of the data show that being from a dominant caste and from a wealthier household is correlated with having higher self-efficacy and self-esteem, as well as better peer-relations.

At endline we also measured girl's resilience and decision making behaviour, particularly the girl's tendency to evaluate their decisions (vigilance) or to avoid making decisions (buck-passing) (see Section 4.2.3). The data shows that being older and being from the dominant caste are correlated with being more resilient or capable of coping with stress, as well as with assessing alternative decisions and avoiding buck passing when making decisions.

6 IMPACT ANALYSIS

6.1 EMPIRICAL STRATEGY

As described in Section 4.1, our design allows us to establish a causal impact of *being eligible for* the PANKH programme (intent-to-treat effect - ITT). In other words, we compare outcomes of girls who live in "treatment" villages (irrespective of whether they participated in *PAnKH*) to those who live in "control" villages.

More formally, for each outcome of interest, we can identify the impact of the Girl Only (T^G) and the Integrated Arm (T^I) as:

$$y_{ij} = \alpha + \beta_G T_j^G + \beta_I T_j^I + \gamma X_{ij} + \varepsilon_{ij} \qquad (1)$$

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 integrated (girls+community) treatment group and equal to zero otherwise. In all inference we will allow for clustering of the random error term ε_{ij} at the cluster level, the unit of randomisation.

To increase the precision of our estimates we will control for key baseline characteristics, X_{ij}. For the outcomes measured at the girl level, we control for a set of girl characteristics (age in years, school attendance, interaction between age and school attendance, caste), household characteristics (wealth index) and carer characteristics (carer education in years). For the outcomes measured at the carer level, we control for a set of carer characteristics (age and caste) and household characteristics (wealth index). In addition to these core controls we control for baseline measures that are closely related to the outcome of interest[14]. 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. This imputation will be accounted for with a dummy variable equal to one for imputed observations.

In this regression framework the most interesting parameters are β_G and β_I, our estimates of the impact of being eligible for the girls' only programme and the integrated programme respectively. It is the size and significance of these parameters that will tell us the impact of the intervention on the outcome of interest and the degree of uncertainty associated with that 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) $H_0: \beta_G = \beta_C$.

For all continuous outcome measures we estimate equation (1) using Ordinary Least Squares (OLS) while in the case of binary outcome variables we will use a logistic regression model. In this later case, we report estimated the difference in proportion of each outcome between treatment groups and the standard error of that difference using STATA's margins command.

We control for multiple hypothesis testing using a stepdown procedure proposed by Romano and Wolf which controls for the probability of falsely rejecting at least one null hypothesis when there is more than one measure of the same outcome, accounting for dependence across tests (Romano & Wolf, 2005).

6.2 MAIN IMPACTS

We analyse the impacts of the girl only and integrated arms of *PAnKH* on three groups of outcomes[15]: behaviours, well-being and attitudes. We study impacts on behaviours related to schooling and marriage; several well-being domains within mental health and life skills; and attitudes relating to gender norms, violence, school, as well as knowledge of sexual and reproductive health. We first report impacts and then follow up with discussion of these in Section 7.

Impacts are estimated following the strategy described in Section 6.1 above and model specified in Equation (1) above. Tables with all of the main results can be found in Appendix A.

---

[14] See Pre-analysis plan (Appendix C)

[15] All outcomes are pre-specified in the pre-analysis plan (Appendix C).

Behaviour

Education

At this stage it is not possible to study impacts of the intervention on final level of schooling attained since many girls (56.5%) were still at school at endline. The best we can do is evaluate the impact on the probability of still being at school, since we have this outcome measure for the whole sample. Table A1 shows that girls in both intervention arms were significantly more likely to be attending school at endline than girls in the control group. The effect sizes across the two arms are very similar – girls in both were just under 4 percentage points more likely to be attending school at endline than girls in the control group. This is equivalent to a 7 percent increase relative to the control mean attendance rate at endline of 54%. The results further suggest that these impacts are driven by girls who were at school during the baseline rather than incentivising girls who had dropped out of school to return to school; effects are larger for a sub-group of girls who were at school at baseline.

Effects appear to be driven by the older girls in the sample. While impacts are not significant among girls who were 12-14 at baseline, there is a 6.5 percentage point increase in the probability of attending school at endline among the 15-17 year old girls in both arms, which is equivalent to a 19 percent increase relative to the control mean.

We further explore the age-specific trend in the coefficient plot in Figure 11 which shows school drop-out rates at end-line by age and treatment status. There is a reduction in dropout in both the treatment groups among girls who were 17, 18 and 19 at the time of the endline. The effect is statistically significant in both intervention arms at ages 17 and 18 and the girl only arm for age 19. By far the biggest effect (and the only one that survives multiple hypothesis testing) is the reduction in the probability of dropping out before or at age 18 in the girl only arm.

Figure 11: Effect of *PAnKH* on school drop-out by age at endline
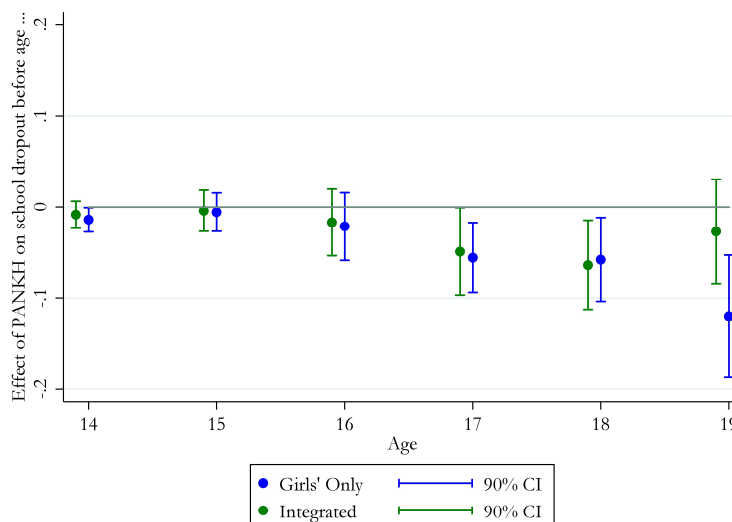
Interestingly, we also see that among the girls who are still at school at endline, there is a significant reduction in the days of school missed in the girl only and integrated arms. On average girls in treatment villages miss 23% less days than girls living in control villages (Table A3). This is a positive sign suggesting that in the medium term we may see the impact of the intervention on schooling grow.

We do not, however, see any impacts of *PAnKH* on attitudes to school (Table A2). The scale used to construct this measure asks about whether girls like school, find it interesting, think it is useful etc.

Marriage

Table A4 shows that overall neither the girl only nor the integrated intervention had a significant impact on the overall probability of being married by the time of the endline. However, as with schooling, there is some suggestive evidence that the intervention had an impact on marriage rates among older girls, age 15-17 at baseline. The proportion of girls in the girl only arm in this age-group who were married by endline is nearly 16 percent lower than in the control group. While the impact in the integrated group is not statistically significant it is only slightly smaller than that in the girl only group and the two are not significantly different from each-other.

Breakdowns by age also show that the marriage effect size increases with age and is highest among girls who are 18-19 at endline; on average there was a 4 percentage point reduction in the proportion of girls 18 or younger who were married by the time of the end-line in the girl only arm (Figure 12). This translates into a 26% decrease relative to the control mean at endline. It is important to note that these effects do not survive multiple hypothesis testing and so are only indicative.

Figure 12: Effect of *PAnKH* on marriage by age at endline
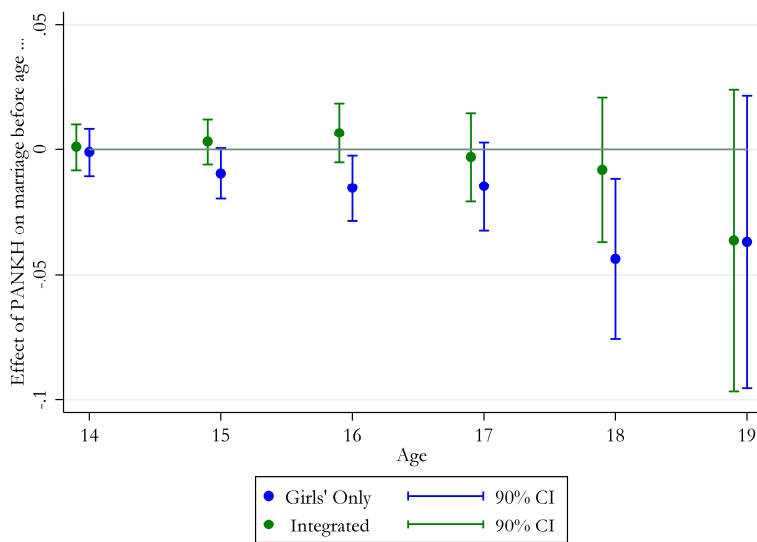
Well-being

Mental Health

A stark finding is that *PAnKH* had clear, highly statistically significant impacts on girls' mental health, but only in the villages where intervention activities included community engagement (Table A5). There is an overall impact of a nearly 0.18 of a standard deviation decrease in an aggregated measure of mental health problems among girls living in villages in the integrated arm and no impact (coefficient size ten times smaller at 0.018) among girls living in the girl only arm. Disaggregating this overall score into constituent depression and anxiety scores, we see that the overall reduction reflects a decrease of 16 and 17% of a standard score, respectively, in each of these. Neither changed significantly among girls living in the girl only treatment villages. We can reject the hypothesis of equality of effects on depression and anxiety in the integrated and girl only arms at the 1% level, even accounting for multiple hypothesis testing. There is a marginally significant negative effect of *PAnKH* on rumination but it does not survive multiple hypothesis testing.

Table A6 further shows effects at cut-off scores on the anxiety and depression scores which are considered meaningful in the medical literature (Gangulyetal, 2013). First, as before, we see that there are no significant impacts in the girl only arm at any cut-off value. We see that the intervention had a significant effect on the rate of moderate depression (depression score greater than or equal to 5) among the girls in the integrated arm. Proportion of girls with moderate depression has gone down by 5 percentage points which is equivalent to a 25% reduction relative to the control group.. Although the relative size of the effect on severe depression (depression score greater than or equal to 10) is bigger at 30%, it is not statistically significant.

Turning to the anxiety measure, we see that the strongest effect is at the cut-off of 8 points. In the literature, individuals with an anxiety score qual to or above 8 points are considered to have moderate anxiety. The intervention reduced the proportion of girls in this group by 2 percentage points which is equivalent to a 29% reduction relative to the control group and is significantly bigger than the effect in the girl only group.

Life skills

We find no evidence of impacts on our measures of life skills including self-efficacy, self-esteem, peer relations, resilience and decision making (Table A7). Girls living in villages that are in either of the treatment arms did not have different levels of these skills at the end of the intervention than girls living in the control villages. We return to this finding in the Discussion below (next section).

Knowledge and Attitudes

In addition to the attitudes directly linked to the behaviours targeted by the programme (e.g. schooling), we study impacts on a broader set of attitudes related to programme content. These include the gender norms of the girls and their primary caregivers, knowledge and attitudes to sexual and reproductive health and attitudes to violence.

Gender

While the PANKH curriculum places a lot of emphasis on helping girls understand the origin of and question the strong patriarchal gender norms that characterise the environment they live in, the results show no indication that the norms of the girls living in intervention villages (or those of their caregivers) became less patriarchal as the result of the programme.

In fact, there appears to be a *positive* and statistically significant impact in the girl only arm, even after adjustment for multiple hypothesis testing. Analysis of item specific responses suggests that this is not an artefact of, for example, the pattern of responses to a small sub-set of items in the gender scale but is consistently evident in the responses to all the items in the scale. This suggests that if anything the attitudes of the girls living in villages in the girl only arm became more patriarchal, a finding which is hard to explain (Table A8).

Sexual and Reproductive Health

There is also no impact on either knowledge of or attitudes towards sexual and reproductive health - another prominent component of the curriculum. The curriculum teaches the girls about the menstruation-cycle, hones in on the origin of some of the cultural norms relating to restrictions during menstruation and provides information on a broader set of sexual and reproductive health issues. We do not find any evidence of an increase in knowledge or change in attitudes on these issues in either the girl only or the integrated arms (Table A9).

Violence

In contrast, there is evidence of a change in attitudes to violence – another prominent topic in the curriculum. The level of passivity towards violence exhibited by the girls in the integrated arm is significantly lower at endline than that in the control group. Among the subs-scales that comprise this measure, there is most evidence of an effect on reducing victim blaming tendencies (Table A11).

Sport

Finally, participation in sport was a novel and key component of the programme. The results show that, as with school attendance, while the programme has an impact on sports behaviour, it does not change attitudes to sport. Tables A12 and A13 shows that, as expected, girls in treatment villages are much more likely (47-58 percent relative to control) to be participating in sports and spend more time on sports activities (52-73 percent increase relative to girls in the control villages). However, we detect no change at all in their attitudes to sports, measured by questions such as "I can be comfortable with my body while playing sport" and "I can compete in sports just as well as boys".

7 DISCUSSION

Our theory of change specifies five main outcomes: education, marriage, mental health, socio-emotional well-being and gender attitudes (Section 3.2). As reported above, we find

that the intervention had significant effects on three of these – education, marriage and mental health.

A key question is *how* the programme achieves these impacts and, from that, what are the essential and non-essential features for impacting each of the outcomes. It has previously been shown that broad-based "life-skill" interventions impact schooling, marriage and employment related outcomes when bundled with "hard" skills training (such as vocational training programmes) and/or financial resources (e.g. Baird et al, 2011, 2014; Bandiera et al, 2017). However, much of the evidence to date contains little to suggests that behavior can be changed through targeting "life-skills" only (e.g. Buchmann et al, 2017). This would seem an especially likely outcome in our study setting given that girls are particularly disempowered and have even less say on key decisions regarding, among others, their schooling, marriage and family planning than do girls in many other developing country settings. The significant impacts that we find are all the more striking in this context.

The main hypotheses in our theory of change for the way in which *PAnKH* may impact the targeted final outcomes include improved knowledge, attitudinal change, increased social support and creation of enabling environments within the household and the community. In what follows, we discuss the main impacts and indications of potential mechanisms coming out of our impact analysis in more depth.

Marriage and education effects

The first point to note about the education and marriage effects is that they seem to set in at the ages when, in the control group, we see the sharpest decline in school attendance and rise in marriage. Figure 13 shows that this happens roughly between the ages of XX and XX, which aligns closely with the timing of impacts, especially evident in Figures 11 and 12 above.

Figure 13: Marriage and school attendance rates by age at endline (control group)

School attendance by age   Marriage rates by age
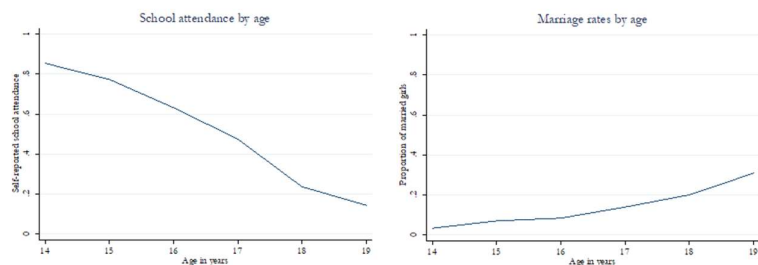
They key question, however, is how could a programme that focuses entirely on girls' well-being, life skills and attitudes change education and marriage outcomes in a setting where girls have so little say in most key decisions made about them. A combination of our baseline data and information from focus group work suggest that a potential mechanism may be girls choosing to stay at school longer, which in turn, results in marriage delays.

Specifically, mothers of adolescent girls in the study district suggested that as long as girls behave "appropriately" and remain interested in school, they can continue going school. Furthermore, that girls were not usually taken out of school to be married, but rather that if an adolescent girl is not in full time education, parents start thinking about her marriage. This is consistent with our baseline data which suggest (a) that girls have a lot more say in the decision about whether they continue going to school than whether they get married; and that (b) behaviours that are most likely to get them taken out of school are within their control.

It is, therefore, plausible that as the result of the intervention, girls could change their behavior in ways that enabled them to stay at school for longer. This is in line with the additional finding that school going girls in intervention villages are less likely to miss school, suggesting that they are more actively involved/interested in school than girls in control villages. This is further consistent with the emphasis in the curriculum on the importance and value of school and evidence in the literature that participation in sport can improve school performance (Stevenson, 2010). Since married girls are highly unlikely to be attending school, staying at school longer could in turn result in a delay in marriage. The intervention could thus affect marriage outcomes indirectly, even in a context where there is nothing girls can do to directly change marriage related decisions.

We test this hypothesis through mediation analysis in which we assess whether school attendance mediates the negative effect of *PAnKH* on the probability of being married at endline. The results are presented in Table A14 and show that the initially statistically

significant effects of *PAnKH* on marriage for the older girls in the girl only arm, goes down in size and is no longer statistically significant once a control for school attendance is added. This is consistent with the hypothesis that the programme affected marriage outcomes through its positive impact on school attendance.

This analysis sheds light on a likely mechanism for the marriage effect but a big question remains about what is driving the school attendance effect. Attitudinal and knowledge change are key components of the *PAnKH* theory of change. The programme is designed to alter girls' attitudes to key life decisions by introducing new ideas and concepts, information, as well as making girls feel less bound by prevailing norms and attitudes. The attitudinal change, in turn, is expected to lead to behavioural change.

However, the programme appears to have had little impact on any of the attitudes or knowledge domains targeted by the curriculum. These include knowledge related to sexual and reproductive health, attitudes to menstruation, gender norms, schooling and sport. This is a striking finding because it leaves us with little evidence on effectiveness of the curriculum followed in the girl education sessions. Taken at face value it suggests that it is plausible that the impacts observed are driven by the increased social support or higher levels of motivation that girls experience as the result of participation in the groups and sports sessions (hard-to-measure mediators that we do not capture in the data).

An alternative interpretation emerges from literature in psychology on acculturation (the process of adaptation to another culture). Here, there is evidence to suggest that, in fact, attitudinal change follows behavioural change rather than the other way around (as is assumed in our theory of change). Specifically, studies of adaptation of immigrants to new cultural settings show that practices "acculturate" more quickly than attitudes. This has been found in parenting behaviour (Bornstein & Cole, 2006), but also more generally (Nagata, 1994). In the context of our study this suggests that the curriculum content could still be affecting behavior even if we do not yet see attitudinal change which is consistent with this.

In addition, measurement of knowledge and attitudes is challenging, especially in new contexts. There is, therefore, a risk that that our measures are simply inadequate for detecting programme effects. Although we cannot rule this out, it is unlikely to be driving the findings in this instance as all of the measures used in the analysis show strong psychometric properties suggesting that they performed well[16].

Mental health effects

We now turn to the striking finding that the addition of the community engagement component to *PAnKH* activities has strong positive effects on girls' mental health (depression and anxiety).

The aim of community engagement was to make the environment in which the girls live more receptive to their needs, open to their input and feel safer for them. It was very clear from

---

[16] Results of analysis of the psychometric properties of the measures is available on request.

social mapping work conducted before the baseline that fear of violence and harassment are a very prominent feature of girls' day-to-day experience in their communities outside their homes. When asked to mark areas on the village map where girls can play and feel safe, men and boys typically suggested that girls would feel safe in the entire village, especially during daylight hours. In contrast, women and girls, generally reported feeling safe in the smallest areas of the village, often confined only to their own homes. Some girls reported not wanting even to walk to the village health centre (*Anganwadi*) alone for fear of violence or harassment. As described in detail in the companion *Programme Implementation and Cost Effectiveness Analysis* report, a core element of the community engagement was monthly Call for Action Events during which the girls communicated material covered in the education sessions to members of their communities, including on issues of violence, harassment, inequitable gender norms etc.

Plausibly, these events may have made harassment of girls in public less acceptable and/or given girls a stronger sense of support within their communities which led to the significant gains in their mental health. Due to resource constraints we were unable to collect the data that would be needed to directly test this hypothesis though results on reduction in passivity of the girls attitude towards violence in the integrated arm only are consistent with girls feeling less vulnerable and disempowered.

Life-skills effects

There is no evidence of effects on life skills outside the mental health effects. As with knowledge and attitudinal measures, it is important to allow for the possibility of inadequate measurement. However, there is little in the data to suggest that this should be a major concern.

A point worth noting is that there is surprisingly little evidence in the literature on impacts of life-skills interventions on socio-emotional/life-skills. Typically, studies measure behavioural and attitudinal outcomes but not life-skills. It may, therefore, be the case that our finding would be replicated in other evaluations of life-skills interventions if they contained as comprehensive a set of direct measures of life-skills as does our study.

8 CONCLUSIONS

This study is among the first to rigorously demonstrate potential of life skills interventions to change key outcomes of adolescent girls living in contexts where girls and women continue to be caught in a vicious cycle of low levels of human capital, low labour force participation rate, low wages, low bargaining power within the household, early marriage and high fertility. The results of this trial suggest that life skills interventions can have significant positive impacts on key outcomes such as schooling, marriage and mental health that have the potential to result in permanent improvements in the life trajectories of the girls and their children. Further, it sheds light on some key mechanisms including the important protective role that schooling plays in contexts where early marriage is prevalent, as well as the critical contribution of the wider community environment to girls' mental health. Finally, it offers a blueprint for scalable and cost-effective programmes for improving outcomes of adolescent girls which aligns closely with the priorities and current approach of the government of India.

Key findings include:

- Girls in both intervention arms were significantly more likely to be attending school at endline than girls in the control group.
- There is suggestive evidence that *PAnKH* had an impact on marriage rates among older girls, age 15-17 at baseline.
- The education and marriage effects set in at the ages when girls in the control group experience the sharpest decline in school attendance and rise in marriage.
- Results of mediation analysis suggest that delays in marriage in the treatment group are explained by girls staying at school longer.
- Adolescent girls experienced significant improvements in their mental health and a reduction in tendency to victim-blaming for experience of violence when the programme targeted the wider community environment in addition to the activities with the girls themselves.
- The programme had no significant impacts on life-skills including self-esteem, self-efficacy, peer relations, resilience and decision-making strategies.
- The programme also had no impacts on knowledge and attitudes to sexual and reproductive health.

We estimate that achieving a 1 percentage point increase in the probability of a girl attending school cost 9.27 USD when achieved through the girl only intervention and 12.98 USD through the integrated intervention. Although it is challenging to compare cost-effectiveness across different types of interventions in different contexts, a tentative analysis suggests PAnKH's cost-effectiveness compares favourably to increasing adolescent girls' school attendance through financial incentives.

Taken together, the experience of implementing the PAnKH programmes, the quantitative evaluation, the qualitative evaluation and the cost-effectiveness analysis suggest that community-based interventions that focus on life skills, attitudes and mental health may be a promising model for policy seeking to improve the education and welfare of adolescent girls. However, the design and implementation of PAnKH reinforced the view that programmes that seek to change 'hard' outcomes such as education and marriage through altering subtle, subjective and difficult to measure constructs such as life skills, attitudes, knowledge and mental health are challenging to implement well and require well trained and highly engaged facilitators.

9 REFERENCES

Achyut, P., Andrew, A., Das, M., Gautam, A., Huepe, M., Krutikova, S., … Verma, R. (2016). Promoting Adolescent Engagement, Knowledge and Health Evaluation of PAnKH : an adolescent girl intervention in Rajasthan, India. Baseline Report.

Adhvaryu, A., Kala, N., & Nyshadham, A. (2018). The Skills to Pay the Bills: Returns to On-the-job Soft Skills Training. Cambridge, MA.

Adoho, F., Chakravarty, S., Korkoyah, D. T., Lundberg, M., & Tasneem, A. (2014). The Impact of an Adolescent Girls Employment Program: The EPAG Project in Liberia (Policy Research Working Paper No. 6832).

Baird, S., McIntosh, C., & Özler, B. (2011). Cash or condition? Evidence from a cash transfer experiment. Quarterly Journal of Economics, 126, 1709–1753.

Barry, M. M., Clarke, A. M., Jenkins, R., & Patel, V. (2013). A systematic review of the effectiveness of mental health promotion interventions for young people in low and middle income countries. BMC Public Health, 13, 835.

Buchmann, N., Field, E., Glennerster, R., Nazneen, S., Pimkina, S., & Sen, I. (2017). Power vs Money: Alternative Approaches to Reducing Child Marriage in Bangladesh, a Randomized Control Trial.

Cappelen, A. W., Charness, G., Ekström, M., Gneezy, U., & Tungodden, B. (2017). Exercise Improves Academic Performance Exercise Improves Academic Performance.

Case, B. A., & Paxson, C. (2013). HIV Risk and Adolescent Behaviors in Africa. American Economic Review: Papers and Proceedings, 103. doi:10.1257/aer.103.3.433

Duflo, E., Dupas, P., & Kremer, M. (2015). Education, HIV, and early fertility: Experimental evidence from Kenya. American Economic Review, 105, 2757–2797.

Eldreth, D., Hardin, M. G., Pavletic, N., & Ernst, M. (2013). Adolescent Transformations of Behavioral and Neural Processes as Potential Targets for Prevention. Prevention Science, 14, 257–266.

Heckman, J. J., & Rubinstein, Y. (2001). The benefits of skill: The importance of noncognitive skills: Lessons from the GED testing program. American Economic Review, 91, 145–154.

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, 24, 411–482.

Jayachandran, S. (2015). The Roots of Gender Inequality in Developing Countries. Annual Review of Economics, 7, 63–88.

Kia-Keating, M., Dowdy, E., Morgan, M. L., & Noam, G. G. (2011). Protecting and promoting: An integrative conceptual model for healthy development of adolescents. Journal of Adolescent Health, 48, 220–228.

Knoll, L. J., Magis-Weinberg, L., Speekenbrink, M., & Blakemore, S. J. (2015). Social Influence on Risk Perception During Adolescence. Psychological Science, 26, 583–592.

Leventhal, K. S., Gillham, J., DeMaria, L., Andrew, G., Peabody, J., & Leventhal, S. (2015). Building psychosocial assets and wellbeing among adolescent girls: A randomized controlled trial. Journal of Adolescence, 45, 284–295.

Romano, J. P., & Wolf, M. (2005). Stepwise Multiple Testing as Formalized Data Snooping. Econometrica, 73, 1237–1282.

Smith-Donald, R., Raver, C. C., Hayes, T., & Richardson, B. (2007). Preliminary construct and concurrent validity of the Preschool Self-regulation Assessment (PSRA) for field-based research. Early Childhood Research Quarterly, 22, 173–187.

Stevenson, B. (2010). Beyond the classroom: Using title IX to measure the return to high school sports. Review of Economics and Statistics, 92, 284–301.

APPENDIX A: RESULTS TABLES

Table A1: School

	Control Mean	Girl Only	AME	P-value [RW P-value]	Integrated	AME	P-value [RW P-value]	Difference P-value [RW P-value]	N
Attending School	0.537	0.281** (0.115)	0.040	0.014	0.271** (0.124)	0.038	0.025	0.936	5703
Attending School (Restricted Sample)	0.562	0.306** (0.125)	0.043	0.014	0.276** (0.137)	0.039	0.038	0.830	5150
Attending School | Attending at BL	0.670	0.310** (0.121)	0.053	0.012	0.264** (0.127)	0.045	0.028	0.731	4348
Attending School (Restricted Sample) | Attending at BL	0.682	0.337*** (0.129)	0.056	0.010	0.272** (0.138)	0.045	0.042	0.662	4000
Attending School | Aged 12-14 at BL	0.758	0.119 (0.151)	0.015	0.422 [0.422]	0.098 (0.164)	0.012	0.545 [0.545]	0.899 [0.990]	2947
Attending School | Aged 15-17 at BL	0.306	0.431*** (0.154)	0.067	0.005 [0.011]	0.417*** (0.148)	0.064	0.002 [0.006]	0.931 [0.990]	2754
Attending School | Aged 12-14 at BL & Attending at BL	0.818	0.160 (0.159)	0.021	0.315 [0.315]	0.101 (0.177)	0.013	0.562 [0.562]	0.746 [0.936]	2619
Attending School | Aged 15-17 at BL & Attending at BL	0.451	0.451*** (0.163)	0.102	0.006 [0.014]	0.402*** (0.148)	0.090	0.004 [0.007]	0.761 [0.936]	1727
Dropped school ≤ 13 years | Attending at BL & Age ≥ 13	0.049	-0.341* (0.198)	-0.014	0.082 [0.233]	-0.204 (0.224)	-0.009	0.350 [0.807]	0.553 [0.925]	4000
Dropped school ≤ 14 years | Attending at BL & Age ≥ 14	0.075	-0.079 (0.194)	-0.005	0.675 [0.675]	-0.056 (0.208)	-0.004	0.797 [0.852]	0.924 [0.995]	4000
Dropped school ≤ 15 years | Attending at BL & Age ≥ 15	0.142	-0.193 (0.208)	-0.021	0.358 [0.647]	-0.155 (0.206)	-0.017	0.450 [0.852]	0.868 [0.995]	2956
Dropped school ≤ 16 years | Attending at BL & Age ≥ 16	0.220	-0.380** (0.159)	-0.056	0.016 [0.072]	-0.333* (0.204)	-0.050	0.087 [0.323]	0.808 [0.994]	2290
Dropped school ≤ 17 years | Attending at BL & Age ≥ 17	0.270	-0.344** (0.165)	-0.060	0.033 [0.128]	-0.370** (0.174)	-0.064	0.031 [0.129]	0.874 [0.995]	1547
Dropped school ≤ 18 years | Attending at BL & Age ≥ 18	0.304	-0.697*** (0.248)	-0.118	0.006 [0.037]	-0.147 (0.198)	-0.027	0.479 [0.852]	0.030 [0.142]	861
Dropped school ≤ 19 years | Attending at BL & Age ≥ 19	0.243	-0.426 (0.424)	-0.063	0.347 [0.647]	0.261 (0.373)	0.044	0.528 [0.852]	0.100 [0.334]	341

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome variable is a dummy that takes the value of 1 if the girl is currently attending school. The specifications include baseline controls: wealth index, caste, block, mother's education in years, full interactions between school attendance and age in years, carer's gender patriarchal attitudes, girl's intended age of marriage (reported by caregiver), marriage fixed or engaged, a dummy that captures if the elders were taking about marriage and a dummy equal to 1 if the girl was married with gauna not performed. In the case of missing baseline controls we imputed the missing covariate value with the average (mean for continuous controls and median for discrete) of the non-missing observations and this imputation was accounted for with a dummy variable equal to one for imputed observations. We allowed for clustering of the random error term at the cluster level. P-values are two-sided and constructed using a t-bootstrap accounting for stratification by region and treatment status. RW P-values use the Romano-Wolf method to account for multiple hypothesis testing across all outcomes measures in table.

1

Table A2: Attitude to School

	Girl Only	P-value [RW P-value]	Integrated	P-value [RW P-value]	Difference P-value [RW P-value]	N
School Attitudes	0.001 (0.082)	0.994 [0.994]	0.013 (0.077)	0.846 [0.846]	0.863 [0.959]	3094
Past-School Attitudes	0.031 (0.080)	0.692 [0.881]	0.047 (0.072)	0.494 [0.692]	0.804 [0.959]	1918

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome variables were measured using exploratory factor analysis for a scale on attitudes towards school. The specifications include baseline controls: wealth index, caste, block, mother's education in years, full interactions between school attendance and age in years, and a dummy equal to 1 if the girl was married with gauna not performed. In the case of missing baseline controls we imputed the missing covariate value with the average (mean for continuous controls and median for discrete) of the non-missing observations and this imputation was accounted for with a dummy variable equal to one for imputed observations. We allowed for clustering of the random error term at the cluster level. P-values are two-sided and constructed using a t-bootstrap accounting for stratification by region and treatment status. RW P-values use the Romano-Wolf method to account for multiple hypothesis testing across all outcomes measures in table.

Table A3: Number of days of school missed

	Control Mean	Girl Only	P-value	Integrated	P-value	N
Days miss school in typical month	3.538	-0.809* (0.452)	0.068	-0.781** (0.415)	0.048	3107

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The specifications include baseline controls: wealth index, caste, block, mother's education in years, full interactions between school attendance and age in years, caregiver's gender attitudes, girl's intended age of marriage (reported by caregiver), marriage fixed or engaged, a dummy that captures if the elders were taking about marriage and a dummy equal to 1 if the girl was married with gauna not performed. We also control for the baseline measures that were closely related to the outcome of interest: carer's gender patriarchal attitudes at baseline. In the case of missing baseline controls we imputed the missing covariate value with the average (mean for continuous controls and median for discrete) of the non-missing observations and this imputation was accounted for with a dummy variable equal to one for imputed observations. We allowed for clustering of the random error term at the cluster level. P-values are two-sided and constructed using a t-bootstrap accounting for stratification by region and treatment status. RW P-values use the Romano-Wolf method to account for multiple hypothesis testing across all outcomes measures in table.

Table A4: Marriage

	Control Mean	Girl Only	AME	P-value [RW P-value]	Integrated	AME	P-value [RW P-value]	Difference P-value [RW P-value]	N
Married	0.126	-0.176 (0.124)	-0.016	0.153	-0.105 (0.125)	-0.010	0.396	0.611	5635
Married | 12-14 at BL	0.045	-0.065 (0.213)	-0.003	0.758 [0.758]	0.194 (0.228)	0.009	0.393 [0.393]	0.193 [0.337]	2918
Married | 15-17 at BL	0.189	-0.254** (0.132)	-0.032	0.048 [0.100]	-0.203 (0.143)	-0.026	0.154 [0.273]	0.736 [0.736]	2715
Girl Married ≤ 14 Years	0.009	-0.053 (0.327)	-0.001	0.878 [0.878]	0.042 (0.326)	0.001	0.894 [0.963]	0.805 [0.967]	2912
Girl Married ≤ 15 Years	0.017	-0.423 (0.294)	-0.008	0.152 [0.457]	0.162 (0.278)	0.004	0.557 [0.950]	0.023 [0.099]	3722
Girl Married ≤ 16 Years	0.035	-0.444* (0.231)	-0.013	0.058 [0.235]	0.211 (0.221)	0.008	0.349 [0.838]	0.003 [0.011]	3658
Girl Married ≤ 17 Years	0.079	-0.233 (0.163)	-0.015	0.151 [0.457]	-0.042 (0.167)	-0.003	0.810 [0.963]	0.311 [0.663]	2707
Girl Married ≤ 18 Years	0.170	-0.379** (0.169)	-0.043	0.023 [0.146]	-0.075 (0.153)	-0.009	0.621 [0.950]	0.089 [0.283]	1665
Girl Married ≤ 19 Years	0.300	-0.226 (0.220)	-0.037	0.307 [0.515]	-0.259 (0.229)	-0.042	0.259 [0.783]	0.876 [0.967]	753

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Girls who were married (with gauna not performed) at baseline are dropped from the analysis sample to assess treatment effects on marriage. 'Marriage' is a dummy that takes the value of 1 if the girl is currently married, married with gauna not performed, divorced or widowed. 'Married-Engaged' includes the girls that were already engaged at endline. The specifications include baseline controls: wealth index, caste, block, mother's education in years, full interactions between school attendance and age in years, carer's gender patriarchal attitudes at baseline, girl's intended age of marriage (reported by caregiver), marriage fixed or engaged, a dummy that captures if the elders were taking about marriage and a dummy equal to 1 if the girl was married with gauna not performed. In the case of missing baseline controls we imputed the missing covariate value with the average (mean for continuous controls and median for discrete) of the non-missing observations and this imputation was accounted for with a dummy variable equal to one for imputed observations. We allowed for clustering of the random error term at the cluster level. P-values are two-sided and constructed using a t-bootstrap accounting for stratification by region and treatment status. RW P-values use the Romano-Wolf method to account for multiple hypothesis testing across all outcomes measures in table.

3

Table A5: Mental Health

	Girl Only	P-value [RW P-value]	Integrated	P-value [RW P-value]	Difference P-value [RW P-value]	N
PHQ-9 Depression	0.018 (0.076)	0.817 [0.823]	-0.160** (0.064)	0.011 [0.021]	0.002 [0.005]	5150
GAD-7 Anxiety	-0.024 (0.059)	0.684 [0.823]	-0.168*** (0.054)	0.002 [0.006]	0.004 [0.009]	5150
Rumination	-0.148* (0.088)	0.085 [0.197]	0.036 (0.104)	0.720 [0.720]	0.069 [0.069]	5113
Poor Mental Health	-0.018 (0.068)	0.793	-0.178*** (0.061)	0.003	0.004	5113

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome variables were measured using exploratory factor analysis on the following scales: PHQ-9 scale to measure depression, GAD-7 scale to measure anxiety and the short version of the Rumination Responses scale. The Mental Health Factor is the first principal factor from exploratory factor analysis on the three indicators of mental health. Higher values of the Mental Health Factor should be interpreted as lower mental mental health. The specifications include baseline controls: wealth index, caste, block, mother's education in years, full interactions between school attendance and age in years, and a dummy equal to 1 if the girl was married with gauna not performed. We also control for mental health at baseline. In the case of ing baseline controls we imputed the missing covariate value with the average (mean for continuous controls and median for discrete) of the non-missing observations and this imputation was accounted for with a dummy variable equal to one for imputed observations. We allowed for clustering of the random error term at the cluster level. P-values are two-sided and constructed using a t-bootstrap accounting for stratification by region and treatment status. RW P-values use the Romano-Wolf method to account for multiple hypothesis testing across all outcomes measures in table.

4

Table A6: Mental Health: Binary Measures

	Control Mean	Girl Only	AME	P-value [RW P-value]	Integrated	AME	P-value [RW P-value]	Difference P-value [RW P-value]	N
Depression [cut-off >=10]	0.042	-0.017 (0.263)	-0.001	0.946 [1.000]	-0.399 (0.231)	-0.013	0.113 [0.221]	0.097 [0.194]	5146
Depression [cut-off >=5]	0.197	0.013 (0.147)	0.002	0.929 [1.000]	-0.353** (0.139)	-0.050	0.013 [0.049]	0.008 [0.030]	5146
Anxiety [cut-off >=10]	0.032	0.102 (0.215)	0.003	0.639 [0.972]	-0.072 (0.177)	-0.002	0.703 [0.703]	0.439 [0.439]	5146
Anxiety [cut-off >=9]	0.052	-0.034 (0.190)	-0.002	0.846 [0.998]	-0.294* (0.155)	-0.013	0.061 [0.160]	0.174 [0.226]	5146
Anxiety [cut-off >=8]	0.072	-0.006 (0.179)	-0.000	0.977 [1.000]	-0.373** (0.164)	-0.021	0.029 [0.090]	0.031 [0.084]	5146
Anxiety [cut-off >=7]	0.099	0.017 (0.160)	0.002	0.922 [1.000]	-0.250* (0.146)	-0.020	0.095 [0.221]	0.075 [0.179]	5146

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome variables are dummy variables that take the value of 1 if the total continuous score of depression ($PHQ-9$) or anxiety ($GAD-7$) is equal or above a specific cut-off value. The cut-off values for PHQ-9 and GAD-7 have been validated in several countries and for different population groups. However, only the PHQ-9 has been validated in a population of adolescent girls in India (*Gangulyetal*, 2013). The specifications include baseline controls: wealth index, caste, block, mother's education in years, full interactions between school attendance and age in years, and a dummy equal to 1 if the girl was married with gauna not performed. We also control for mental health at baseline. In the case of missing baseline controls we imputed the missing covariate value with the average (mean for continuous controls and median for discrete) of the non-missing observations and this imputation was accounted for with a dummy variable equal to one for imputed observations. We allowed for clustering of the random error term at the cluster level. P-values are two-sided and constructed using a t-bootstrap accounting for stratification by region and treatment status. RW P-values use the Romano-Wolf method to account for multiple hypothesis testing across all outcomes measures in table.

Table A7: Life Skills

	Girl Only	P-value [RW P-value]	Integrated	P-value [RW P-value]	Difference P-value [RW P-value]	N
Self-efficacy	-0.028 (0.089)	0.751 [0.949]	0.017 (0.098)	0.863 [0.987]	0.599 [0.782]	5133
Self-esteem	0.003 (0.091)	0.973 [0.996]	0.071 (0.090)	0.402 [0.765]	0.399 [0.761]	5117
Peer relations	0.005 (0.081)	0.949 [0.996]	-0.014 (0.075)	0.861 [0.987]	0.793 [0.793]	5132
Resilience	-0.135* (0.082)	0.091 [0.297]	-0.017 (0.095)	0.854 [0.987]	0.146 [0.403]	5105
Decision making- vigilance	-0.089 (0.065)	0.155 [0.444]	0.033 (0.080)	0.702 [0.963]	0.081 [0.300]	5119
Decision making - buck passing	0.031 (0.066)	0.640 [0.949]	0.078 (0.067)	0.247 [0.626]	0.389 [0.761]	5121
Life Skills Factor	-0.027 (0.093)	0.765	0.026 (0.100)	0.795	0.541	5021

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome variables were measured using exploratory factor analysis on the following scales: General Self-Efficacy scale, sub-scales of the Self-Description Questionnaire to measure Self-esteem and Peer Relations, Connor-Davidson Resilience Scale and the Melbourne Decision Making Questionnaire to measure Vigilance and Buck-passing. The Soft Skills factor is the first principal factor from exploratory factor analysis on all soft-skill indicators. The specifications include baseline controls: wealth index, caste, block, mother's education in years, full interactions between school attendance and age in years, and a dummy equal to 1 if the girl was married with gauna not performed. We also control for the baseline measures that were closely related to the outcome of interest: self-efficacy, self-esteem, peers relations and socio-emotional skills. In the case of missing baseline controls we imputed the missing covariate value with the average (mean for continuous controls and median for discrete) of the non-missing observations and this imputation was accounted for with a dummy variable equal to one for imputed observations. We allowed for clustering of the random error term at the cluster level. P-values are two-sided and constructed using a t-bootstrap accounting for stratification by region and treatment status. RW P-values use the Romano-Wolf method to account for multiple hypothesis testing across all outcomes measures in table.

Table A8: Gender Attitudes

	Girl Only	P-value [RW P-value]	Integrated	P-value [RW P-value]	Difference P-value [RW P-value]	N
Girl patriarchal gender attitudes	0.113** (0.056)	0.047 [0.084]	-0.079 (0.065)	0.205 [0.338]	0.002 [0.005]	5088
Carer patriarchal gender attitudes	0.068 (0.066)	0.289 [0.289]	-0.040 (0.072)	0.577 [0.577]	0.119 [0.119]	4846
Gender Attitudes Index (Average)	0.099 (0.064)	0.113	-0.066 (0.073)	0.354	0.017	4791

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome variables were measured using exploratory factor analysis for the adapted Gender Equitable Men Scale. The specifications include baseline controls: wealth index, caste, block, mother's education in years, mother's age and age squared, full interactions between school attendance and age in years, and a dummy equal to 1 if the girl was married with gauna not performed. We also control for girl and carer gender patriarchal attitudes at baseline. In the case of missing baseline controls we imputed the missing covariate value with the average (mean for continuous controls and median for discrete) of the non-missing observations and this imputation was accounted for with a dummy variable equal to one for imputed observations. We allowed for clustering of the random error term at the cluster level. P-values are two-sided and constructed using a t-bootstrap accounting for stratification by region and treatment status. RW P-values use the Romano-Wolf method to account for multiple hypothesis testing across all outcomes measures in table.

Table A9: Knowledge of SRH

	Girl Only	P-value [RW P-value]	Integrated	P-value [RW P-value]	Difference P-value [RW P-value]	N
Knowledge Menstruation-Girls	-0.066 (0.062)	0.287 [0.939]	0.008 (0.059)	0.885 [0.506]	0.207 [0.520]	4982
Knowledge SRH-Girls	0.048 (0.062)	0.423 [0.567]	0.070 (0.062)	0.238 [0.450]	0.739 [0.739]	4154
Knowledge Index (Average)	-0.005 (0.068)	0.939	0.051 (0.058)	0.340	0.400	4122

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome variables were measured using Item Response Theory for a puberty and menstruation knowledge scale, and a contraception and sexual knowledge scale. The specifications include baseline controls: wealth index, caste, block, mother's education in years, full interactions between school attendance and age in years, and a dummy equal to 1 if the girl was married with gauna not performed. We also control for the knowledge of menstruation and contraceptive knowledge at baseline. In the case of missing baseline controls we imputed the missing covariate value with the average (mean for continuous controls and median for discrete) of the non-missing observations and this imputation was accounted for with a dummy variable equal to one for imputed observations. We allowed for clustering of the random error term at the cluster level. P-values are two-sided and constructed using a t-bootstrap accounting for stratification by region and treatment status. RW P-values use the Romano-Wolf method to account for multiple hypothesis testing across all outcomes measures in table.

Table A10: Restrictions during Menstruation

	Girl Only	P-value [RW P-value]	Integrated	P-value [RW P-value]	Difference P-value [RW P-value]	N
Restriction Menstruation-Girls	0.080 (0.076)	0.276	0.011 (0.081)	0.885	0.478	4700

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome variable was measured using Item Response Theory for a 'restrictions during menstruation' scale. The specification includes baseline controls: wealth index, caste, block, mother's education in years, full interactions between school attendance and age in years, and a dummy equal to 1 if the girl was married with gauna not performed. We also control for restrictions during menstruation at baseline. In the case of missing baseline controls we imputed the missing covariate value with the average (mean for continuous controls and median for discrete) of the non-missing observations and this imputation was accounted for with a dummy variable equal to one for imputed observations. We allowed for clustering of the random error term at the cluster level. P-values are two-sided and constructed using a t-bootstrap accounting for stratification by region and treatment status. RW P-values use the Romano-Wolf method to account for multiple hypothesis testing across all outcomes measures in table.

Table A11: Violence Attitudes and Responses

	Girl Only	P-value [RW P-value]	Integrated	P-value [RW P-value]	Difference P-value [RW P-value]	N
Victim blaming tendencies	-0.034 (0.055)	0.547 [0.757]	-0.105** (0.048)	0.023 [0.107]	0.144 [0.415]	5150
Perpetrator blaming tendencies	-0.010 (0.061)	0.873 [0.873]	0.034 (0.061)	0.571 [0.958]	0.507 [0.618]	5149
Avoidance behaviours tendencies	0.061 (0.050)	0.230 [0.501]	0.013 (0.052)	0.811 [0.992]	0.380 [0.618]	5149
Retaliation behaviours tendencies	-0.096 (0.067)	0.155 [0.441]	-0.009 (0.065)	0.883 [0.992]	0.251 [0.541]	5149
Reporting behaviours tendencies	-0.120** (0.056)	0.030 [0.127]	0.011 (0.052)	0.818 [0.992]	0.036 [0.140]	5149
Violence Index: passive attitudes|responses to violence	-0.010 (0.053)	0.850	-0.086* (0.045)	0.059	0.127	5149

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome variables were measured using Item Response Theory on a violence scale. The Violence Index is the first principal factor from exploratory factor analysis on the indicators of violence: victimization and avoidance behaviour loaded positively; and perpetrator, retaliation and reporting behaviour loaded negatively. Higher values of the Violence Index should be interpreted as a higher acceptance of violence or more passive attitude\|responses to violence. The specifications include baseline controls: wealth index, caste, block, mother's education in years, full interactions between school attendance and age in years, and a dummy equal to 1 if the girl was married with gauna not performed. We also control for the girl's gender patriarchal attitudes at baseline. In the case of missing baseline controls we imputed the missing covariate value with the average (mean for continuous controls and median for discrete) of the non-missing observations and this imputation was accounted for with a dummy variable equal to one for imputed observations. We allowed for clustering of the random error term at the cluster level. P-values are two-sided and constructed using a t-bootstrap accounting for stratification by region and treatment status. RW P-values use the Romano-Wolf method to account for multiple hypothesis testing across all outcomes measures in table.

Table A12: Attitudes to and Participation in Sports

	Control Mean	Girl Only	P-value [RW P-value]	Integrated	P-value [RW P-value]	Difference P-value [RW P-value]	N
Sports Attitudes	-0.000	0.007 (0.067)	0.921 [0.921]	0.049 (0.072)	0.488 [0.488]	0.528 [0.528]	5126
Time (minutes) playing sports per day	15.265	11.215*** (2.620)	0.000 [0.000]	7.972*** (2.174)	0.000 [0.001]	0.280 [0.461]	5150

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. 'Sport attitudes' was measured using exploratory factor analysis on a sport attitudes scale. 'Time playing sports' was measured as the total time spent playing sports in a typical week (minutes). The specifications include baseline controls: wealth index, caste, block, mother's education in years, full interactions between school attendance and age in years, and a dummy equal to 1 if the girl was married with gauna not performed. We also control for the baseline measures that were closely related to the outcome of interest: frequency playing sports and sports attitudes scale at baseline. In the case of missing baseline controls we imputed the missing covariate value with the average (mean for continuous controls and median for discrete) of the non-missing observations and this imputation was accounted for with a dummy variable equal to one for imputed observations. We allowed for clustering of the random error term at the cluster level. P-values are two-sided and constructed using a t-bootstrap accounting for stratification by region and treatment status. RW P-values use the Romano-Wolf method to account for multiple hypothesis testing across all outcomes measures in table.

Table A13: Participation in Sports

	Control Mean	Girl Only	AME	P-value	Integrated	AME	P-value	N
Girl participates in sports	0.156	0.621*** (0.131)	0.091	0.000	0.523*** (0.129)	0.074	0.001	5146

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. P-values are two-sided and constructed using a t-bootstrap accounting for stratification by region and treatment status. RW P-values use the Romano-Wolf method to account for multiple hypothesis testing across all outcomes measures in table.

Table A14: Mediation Analysis

	(1) Attending School	(2) Married | 15-17 at BL	(3) Married | 15-17 at BL
Girl only arm treatment	0.419*** (0.154)	-0.254* (0.132)	-0.171 (0.134)
Integrated arm treatment	0.408*** (0.147)	-0.203 (0.143)	-0.140 (0.143)
Attending School			-1.541*** (0.211)
Observations	2715	2715	2715
P-value: diff. in Girl only coef			0.00517
P-value: diff. in Integrated coef			0.0337

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. P-values are two-sided and constructed analytically, accounting for clustering.

APPENDIX A: RESULTS TABLES

Table A1: School

	Control Mean	Girl Only	AME	P-value [RW P-value]	Integrated	AME	P-value [RW P-value]	Difference P-value [RW P-value]	N
Attending School	0.537	0.281** (0.115)	0.040	0.014	0.271** (0.124)	0.038	0.025	0.936	5703
Attending School (Restricted Sample)	0.562	0.306** (0.125)	0.043	0.014	0.276** (0.137)	0.039	0.038	0.830	5150
Attending School | Attending at BL	0.670	0.310** (0.121)	0.053	0.012	0.264** (0.127)	0.045	0.028	0.731	4348
Attending School (Restricted Sample) | Attending at BL	0.682	0.337*** (0.129)	0.056	0.010	0.272** (0.138)	0.045	0.042	0.662	4000
Attending School | Aged 12-14 at BL	0.758	0.119 (0.151)	0.015	0.422 [0.422]	0.098 (0.164)	0.012	0.545 [0.545]	0.899 [0.990]	2947
Attending School | Aged 15-17 at BL	0.306	0.431*** (0.154)	0.067	0.005 [0.011]	0.417*** (0.148)	0.064	0.002 [0.006]	0.931 [0.990]	2754
Attending School | Aged 12-14 at BL & Attending at BL	0.818	0.160 (0.159)	0.021	0.315 [0.315]	0.101 (0.177)	0.013	0.562 [0.562]	0.746 [0.936]	2619
Attending School | Aged 15-17 at BL & Attending at BL	0.451	0.451*** (0.163)	0.102	0.006 [0.014]	0.402*** (0.148)	0.090	0.004 [0.007]	0.761 [0.936]	1727
Dropped school ≤ 13 years | Attending at BL & Age ≥ 13	0.049	-0.341* (0.198)	-0.014	0.082 [0.233]	-0.204 (0.224)	-0.009	0.350 [0.807]	0.553 [0.925]	4000
Dropped school ≤ 14 years | Attending at BL & Age ≥ 14	0.075	-0.079 (0.194)	-0.005	0.675 [0.675]	-0.056 (0.208)	-0.004	0.797 [0.852]	0.924 [0.995]	4000
Dropped school ≤ 15 years | Attending at BL & Age ≥ 15	0.142	-0.193 (0.208)	-0.021	0.358 [0.647]	-0.155 (0.206)	-0.017	0.450 [0.852]	0.868 [0.995]	2956
Dropped school ≤ 16 years | Attending at BL & Age ≥ 16	0.220	-0.380** (0.159)	-0.056	0.016 [0.072]	-0.333* (0.204)	-0.050	0.087 [0.323]	0.808 [0.994]	2290
Dropped school ≤ 17 years | Attending at BL & Age ≥ 17	0.270	-0.344** (0.165)	-0.060	0.033 [0.128]	-0.370** (0.174)	-0.064	0.031 [0.129]	0.874 [0.995]	1547
Dropped school ≤ 18 years | Attending at BL & Age ≥ 18	0.304	-0.697*** (0.248)	-0.118	0.006 [0.037]	-0.147 (0.198)	-0.027	0.479 [0.852]	0.030 [0.142]	861
Dropped school ≤ 19 years | Attending at BL & Age ≥ 19	0.243	-0.426 (0.424)	-0.063	0.347 [0.647]	0.261 (0.373)	0.044	0.528 [0.852]	0.100 [0.334]	341

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome variable is a dummy that takes the value of 1 if the girl is currently attending school. The specifications include baseline controls: wealth index, caste, block, mother's education in years, full interactions between school attendance and age in years, carer's gender patriarchal attitudes, girl's intended age of marriage (reported by caregiver), marriage fixed or engaged, a dummy that captures if the elders were taking about marriage and a dummy equal to 1 if the girl was married with gauna not performed. In the case of missing baseline controls we imputed the missing covariate value with the average (mean for continuous controls and median for discrete) of the non-missing observations and this imputation was accounted for with a dummy variable equal to one for imputed observations. We allowed for clustering of the random error term at the cluster level. P-values are two-sided and constructed using a t-bootstrap accounting for stratification by region and treatment status. RW P-values use the Romano-Wolf method to account for multiple hypothesis testing across all outcomes measures in table.

1

Table A2: Attitude to School

	Girl Only	P-value [RW P-value]	Integrated	P-value [RW P-value]	Difference P-value [RW P-value]	N
School Attitudes	0.001 (0.082)	0.994 [0.994]	0.013 (0.077)	0.846 [0.846]	0.863 [0.959]	3094
Past-School Attitudes	0.031 (0.080)	0.692 [0.881]	0.047 (0.072)	0.494 [0.692]	0.804 [0.959]	1918

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome variables were measured using exploratory factor analysis for a scale on attitudes towards school. The specifications include baseline controls: wealth index, caste, block, mother's education in years, full interactions between school attendance and age in years, and a dummy equal to 1 if the girl was married with gauna not performed. In the case of missing baseline controls we imputed the missing covariate value with the average (mean for continuous controls and median for discrete) of the non-missing observations and this imputation was accounted for with a dummy variable equal to one for imputed observations. We allowed for clustering of the random error term at the cluster level. P-values are two-sided and constructed using a t-bootstrap accounting for stratification by region and treatment status. RW P-values use the Romano-Wolf method to account for multiple hypothesis testing across all outcomes measures in table.

Table A3: Number of days of school missed

	Control Mean	Girl Only	P-value	Integrated	P-value	N
Days miss school in typical month	3.538	-0.809* (0.452)	0.068	-0.781** (0.415)	0.048	3107

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The specifications include baseline controls: wealth index, caste, block, mother's education in years, full interactions between school attendance and age in years, caregiver's gender attitudes, girl's intended age of marriage (reported by caregiver), marriage fixed or engaged, a dummy that captures if the elders were taking about marriage and a dummy equal to 1 if the girl was married with gauna not performed. We also control for the baseline measures that were closely related to the outcome of interest: carer's gender patriarchal attitudes at baseline. In the case of missing baseline controls we imputed the missing covariate value with the average (mean for continuous controls and median for discrete) of the non-missing observations and this imputation was accounted for with a dummy variable equal to one for imputed observations. We allowed for clustering of the random error term at the cluster level. P-values are two-sided and constructed using a t-bootstrap accounting for stratification by region and treatment status. RW P-values use the Romano-Wolf method to account for multiple hypothesis testing across all outcomes measures in table.

2

Table A4: Marriage

	Control Mean	Girl Only	AME	P-value [RW P-value]	Integrated	AME	P-value [RW P-value]	Difference P-value [RW P-value]	N
Married	0.126	-0.176 (0.124)	-0.016	0.153	-0.105 (0.125)	-0.010	0.396	0.611	5635
Married | 12-14 at BL	0.045	-0.065 (0.213)	-0.003	0.758 [0.758]	0.194 (0.228)	0.009	0.393 [0.393]	0.193 [0.337]	2918
Married | 15-17 at BL	0.189	-0.254** (0.132)	-0.032	0.048 [0.100]	-0.203 (0.143)	-0.026	0.154 [0.273]	0.736 [0.736]	2715
Girl Married ≤ 14 Years	0.009	-0.053 (0.327)	-0.001	0.878 [0.878]	0.042 (0.326)	0.001	0.894 [0.963]	0.805 [0.967]	2912
Girl Married ≤ 15 Years	0.017	-0.423 (0.294)	-0.008	0.152 [0.457]	0.162 (0.278)	0.004	0.557 [0.950]	0.023 [0.099]	3722
Girl Married ≤ 16 Years	0.035	-0.444* (0.231)	-0.013	0.058 [0.235]	0.211 (0.221)	0.008	0.349 [0.838]	0.003 [0.011]	3658
Girl Married ≤ 17 Years	0.079	-0.233 (0.163)	-0.015	0.151 [0.457]	-0.042 (0.167)	-0.003	0.810 [0.963]	0.311 [0.663]	2707
Girl Married ≤ 18 Years	0.170	-0.379** (0.169)	-0.043	0.023 [0.146]	-0.075 (0.153)	-0.009	0.621 [0.950]	0.089 [0.283]	1665
Girl Married ≤ 19 Years	0.300	-0.226 (0.220)	-0.037	0.307 [0.515]	-0.259 (0.229)	-0.042	0.259 [0.783]	0.876 [0.967]	753

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Girls who were married (with gauna not performed) at baseline are dropped from the analysis sample to assess treatment effects on marriage. 'Marriage' is a dummy that takes the value of 1 if the girl is currently married, married with gauna not performed, divorced or widowed. 'Married-Engaged' includes the girls that were already engaged at endline. The specifications include baseline controls: wealth index, caste, block, mother's education in years, full interactions between school attendance and age in years, carer's gender patriarchal attitudes at baseline, girl's intended age of marriage (reported by caregiver), marriage fixed or engaged, a dummy that captures if the elders were taking about marriage and a dummy equal to 1 if the girl was married with gauna not performed. In the case of missing baseline controls we imputed the missing covariate value with the average (mean for continuous controls and median for discrete) of the non-missing observations and this imputation was accounted for with a dummy variable equal to one for imputed observations. We allowed for clustering of the random error term at the cluster level. P-values are two-sided and constructed using a t-bootstrap accounting for stratification by region and treatment status. RW P-values use the Romano-Wolf method to account for multiple hypothesis testing across all outcomes measures in table.

3

Table A5: Mental Health

	Girl Only	P-value [RW P-value]	Integrated	P-value [RW P-value]	Difference P-value [RW P-value]	N
PHQ-9 Depression	0.018 (0.076)	0.817 [0.823]	-0.160** (0.064)	0.011 [0.021]	0.002 [0.005]	5150
GAD-7 Anxiety	-0.024 (0.059)	0.684 [0.823]	-0.168*** (0.054)	0.002 [0.006]	0.004 [0.009]	5150
Rumination	-0.148* (0.088)	0.085 [0.197]	0.036 (0.104)	0.720 [0.720]	0.069 [0.069]	5113
Poor Mental Health	-0.018 (0.068)	0.793	-0.178*** (0.061)	0.003	0.004	5113

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome variables were measured using exploratory factor analysis on the following scales: PHQ-9 scale to measure depression, GAD-7 scale to measure anxiety and the short version of the Rumination Responses scale. The Mental Health Factor is the first principal factor from exploratory factor analysis on the three indicators of mental health. Higher values of the Mental Health Factor should be interpreted as lower mental mental health. The specifications include baseline controls: wealth index, caste, block, mother's education in years, full interactions between school attendance and age in years, and a dummy equal to 1 if the girl was married with gauna not performed. We also control for mental health at baseline. In the case of ing baseline controls we imputed the missing covariate value with the average (mean for continuous controls and median for discrete) of the non-missing observations and this imputation was accounted for with a dummy variable equal to one for imputed observations. We allowed for clustering of the random error term at the cluster level. P-values are two-sided and constructed using a t-bootstrap accounting for stratification by region and treatment status. RW P-values use the Romano-Wolf method to account for multiple hypothesis testing across all outcomes measures in table.

Table A6: Mental Health: Binary Measures

	Control Mean	Girl Only	AME	P-value [RW P-value]	Integrated	AME	P-value [RW P-value]	Difference P-value [RW P-value]	N
Depression [cut-off >=10]	0.042	-0.017 (0.263)	-0.001	0.946 [1.000]	-0.399 (0.231)	-0.013	0.113 [0.221]	0.097 [0.194]	5146
Depression [cut-off >=5]	0.197	0.013 (0.147)	0.002	0.929 [1.000]	-0.353** (0.139)	-0.050	0.013 [0.049]	0.008 [0.030]	5146
Anxiety [cut-off >=10]	0.032	0.102 (0.215)	0.003	0.639 [0.972]	-0.072 (0.177)	-0.002	0.703 [0.703]	0.439 [0.439]	5146
Anxiety [cut-off >=9]	0.052	-0.034 (0.190)	-0.002	0.846 [0.998]	-0.294* (0.155)	-0.013	0.061 [0.160]	0.174 [0.226]	5146
Anxiety [cut-off >=8]	0.072	-0.006 (0.179)	-0.000	0.977 [1.000]	-0.373** (0.164)	-0.021	0.029 [0.090]	0.031 [0.084]	5146
Anxiety [cut-off >=7]	0.099	0.017 (0.160)	0.002	0.922 [1.000]	-0.250* (0.146)	-0.020	0.095 [0.221]	0.075 [0.179]	5146

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome variables are dummy variables that take the value of 1 if the total continuous score of depression ($PHQ - 9$) or anxiety ($GAD - 7$) is equal or above a specific cut-off value. The cut-off values for PHQ-9 and GAD-7 have been validated in several countries and for different population groups. However, only the PHQ-9 has been validated in a population of adolescent girls in India (*Gangulyetal*, 2013). The specifications include baseline controls: wealth index, caste, block, mother's education in years, full interactions between school attendance and age in years, and a dummy equal to 1 if the girl was married with gauna not performed. We also control for mental health at baseline. In the case of missing baseline controls we imputed the missing covariate value with the average (mean for continuous controls and median for discrete) of the non-missing observations and this imputation was accounted for with a dummy variable equal to one for imputed observations. We allowed for clustering of the random error term at the cluster level. P-values are two-sided and constructed using a t-bootstrap accounting for stratification by region and treatment status. RW P-values use the Romano-Wolf method to account for multiple hypothesis testing across all outcomes measures in table.

Table A7: Life Skills

	Girl Only	P-value [RW P-value]	Integrated	P-value [RW P-value]	Difference P-value [RW P-value]	N
Self-efficacy	-0.028 (0.089)	0.751 [0.949]	0.017 (0.098)	0.863 [0.987]	0.599 [0.782]	5133
Self-esteem	0.003 (0.091)	0.973 [0.996]	0.071 (0.090)	0.402 [0.765]	0.399 [0.761]	5117
Peer relations	0.005 (0.081)	0.949 [0.996]	-0.014 (0.075)	0.861 [0.987]	0.793 [0.793]	5132
Resilience	-0.135* (0.082)	0.091 [0.297]	-0.017 (0.095)	0.854 [0.987]	0.146 [0.403]	5105
Decision making- vigilance	-0.089 (0.065)	0.155 [0.444]	0.033 (0.080)	0.702 [0.963]	0.081 [0.300]	5119
Decision making - buck passing	0.031 (0.066)	0.640 [0.949]	0.078 (0.067)	0.247 [0.626]	0.389 [0.761]	5121
Life Skills Factor	-0.027 (0.093)	0.765	0.026 (0.100)	0.795	0.541	5021

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome variables were measured using exploratory factor analysis on the following scales: General Self-Efficacy scale, sub-scales of the Self-Description Questionnaire to measure Self-esteem and Peer Relations, Connor-Davidson Resilience Scale and the Melbourne Decision Making Questionnaire to measure Vigilance and Buck-passing. The Soft Skills factor is the first principal factor from exploratory factor analysis on all soft-skill indicators. The specifications include baseline controls: wealth index, caste, block, mother's education in years, full interactions between school attendance and age in years, and a dummy equal to 1 if the girl was married with gauna not performed. We also control for the baseline measures that were closely related to the outcome of interest: self-efficacy, self-esteem, peers relations and socio-emotional skills. In the case of missing baseline controls we imputed the missing covariate value with the average (mean for continuous controls and median for discrete) of the non-missing observations and this imputation was accounted for with a dummy variable equal to one for imputed observations. We allowed for clustering of the random error term at the cluster level. P-values are two-sided and constructed using a t-bootstrap accounting for stratification by region and treatment status. RW P-values use the Romano-Wolf method to account for multiple hypothesis testing across all outcomes measures in table.

Table A8: Gender Attitudes

	Girl Only	P-value [RW P-value]	Integrated	P-value [RW P-value]	Difference P-value [RW P-value]	N
Girl patriarchal gender attitudes	0.113** (0.056)	0.047 [0.084]	-0.079 (0.065)	0.205 [0.338]	0.002 [0.005]	5088
Carer patriarchal gender attitudes	0.068 (0.066)	0.289 [0.289]	-0.040 (0.072)	0.577 [0.577]	0.119 [0.119]	4846
Gender Attitudes Index (Average)	0.099 (0.064)	0.113	-0.066 (0.073)	0.354	0.017	4791

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome variables were measured using exploratory factor analysis for the adapted Gender Equitable Men Scale. The specifications include baseline controls: wealth index, caste, block, mother's education in years, mother's age and age squared, full interactions between school attendance and age in years, and a dummy equal to 1 if the girl was married with gauna not performed. We also control for girl and carer gender patriarchal attitudes at baseline. In the case of missing baseline controls we imputed the missing covariate value with the average (mean for continuous controls and median for discrete) of the non-missing observations and this imputation was accounted for with a dummy variable equal to one for imputed observations. We allowed for clustering of the random error term at the cluster level. P-values are two-sided and constructed using a t-bootstrap accounting for stratification by region and treatment status. RW P-values use the Romano-Wolf method to account for multiple hypothesis testing across all outcomes measures in table.

Table A9: Knowledge of SRH

	Girl Only	P-value [RW P-value]	Integrated	P-value [RW P-value]	Difference P-value [RW P-value]	N
Knowledge Menstruation-Girls	-0.066 (0.062)	0.287 [0.939]	0.008 (0.059)	0.885 [0.506]	0.207 [0.520]	4982
Knowledge SRH-Girls	0.048 (0.062)	0.423 [0.567]	0.070 (0.062)	0.238 [0.450]	0.739 [0.739]	4154
Knowledge Index (Average)	-0.005 (0.068)	0.939	0.051 (0.058)	0.340	0.400	4122

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome variables were measured using Item Response Theory for a puberty and menstruation knowledge scale, and a contraception and sexual knowledge scale. The specifications include baseline controls: wealth index, caste, block, mother's education in years, full interactions between school attendance and age in years, and a dummy equal to 1 if the girl was married with gauna not performed. We also control for the knowledge of menstruation and contraceptive knowledge at baseline. In the case of missing baseline controls we imputed the missing covariate value with the average (mean for continuous controls and median for discrete) of the non-missing observations and this imputation was accounted for with a dummy variable equal to one for imputed observations. We allowed for clustering of the random error term at the cluster level. P-values are two-sided and constructed using a t-bootstrap accounting for stratification by region and treatment status. RW P-values use the Romano-Wolf method to account for multiple hypothesis testing across all outcomes measures in table.

Table A10: Restrictions during Menstruation

	Girl Only	P-value [RW P-value]	Integrated	P-value [RW P-value]	Difference P-value [RW P-value]	N
Restriction Menstruation-Girls	0.080 (0.076)	0.276	0.011 (0.081)	0.885	0.478	4700

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome variable was measured using Item Response Theory for a 'restrictions during menstruation' scale. The specification includes baseline controls: wealth index, caste, block, mother's education in years, full interactions between school attendance and age in years, and a dummy equal to 1 if the girl was married with gauna not performed. We also control for restrictions during menstruation at baseline. In the case of missing baseline controls we imputed the missing covariate value with the average (mean for continuous controls and median for discrete) of the non-missing observations and this imputation was accounted for with a dummy variable equal to one for imputed observations. We allowed for clustering of the random error term at the cluster level. P-values are two-sided and constructed using a t-bootstrap accounting for stratification by region and treatment status. RW P-values use the Romano-Wolf method to account for multiple hypothesis testing across all outcomes measures in table.

Table A11: Violence Attitudes and Responses

	Girl Only	P-value [RW P-value]	Integrated	P-value [RW P-value]	Difference P-value [RW P-value]	N
Victim blaming tendencies	-0.034 (0.055)	0.547 [0.757]	-0.105** (0.048)	0.023 [0.107]	0.144 [0.415]	5150
Perpetrator blaming tendencies	-0.010 (0.061)	0.873 [0.873]	0.034 (0.061)	0.571 [0.958]	0.507 [0.618]	5149
Avoidance behaviours tendencies	0.061 (0.050)	0.230 [0.501]	0.013 (0.052)	0.811 [0.992]	0.380 [0.618]	5149
Retaliation behaviours tendencies	-0.096 (0.067)	0.155 [0.441]	-0.009 (0.065)	0.883 [0.992]	0.251 [0.541]	5149
Reporting behaviours tendencies	-0.120** (0.056)	0.030 [0.127]	0.011 (0.052)	0.818 [0.992]	0.036 [0.140]	5149
Violence Index: passive attitudes|responses to violence	-0.010 (0.053)	0.850	-0.086* (0.045)	0.059	0.127	5149

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome variables were measured using Item Response Theory on a violence scale. The Violence Index is the first principal factor from exploratory factor analysis on the indicators of violence: victimization and avoidance behaviour loaded positively; and perpetrator, retaliation and reporting behaviour loaded negatively. Higher values of the Violence Index should be interpreted as a higher acceptance of violence or more passive attitude\|responses to violence. The specifications include baseline controls: wealth index, caste, block, mother's education in years, full interactions between school attendance and age in years, and a dummy equal to 1 if the girl was married with gauna not performed. We also control for the girl's gender patriarchal attitudes at baseline. In the case of missing baseline controls we imputed the missing covariate value with the average (mean for continuous controls and median for discrete) of the non-missing observations and this imputation was accounted for with a dummy variable equal to one for imputed observations. We allowed for clustering of the random error term at the cluster level. P-values are two-sided and constructed using a t-bootstrap accounting for stratification by region and treatment status. RW P-values use the Romano-Wolf method to account for multiple hypothesis testing across all outcomes measures in table.

Table A12: Attitudes to and Participation in Sports

	Control Mean	Girl Only	P-value [RW P-value]	Integrated	P-value [RW P-value]	Difference P-value [RW P-value]	N
Sports Attitudes	-0.000	0.007 (0.067)	0.921 [0.921]	0.049 (0.072)	0.488 [0.488]	0.528 [0.528]	5126
Time (minutes) playing sports per day	15.265	11.215*** (2.620)	0.000 [0.000]	7.972*** (2.174)	0.000 [0.001]	0.280 [0.461]	5150

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. 'Sport attitudes' was measured using exploratory factor analysis on a sport attitudes scale. 'Time playing sports' was measured as the total time spent playing sports in a typical week (minutes). The specifications include baseline controls: wealth index, caste, block, mother's education in years, full interactions between school attendance and age in years, and a dummy equal to 1 if the girl was married with gauna not performed. We also control for the baseline measures that were closely related to the outcome of interest: frequency playing sports and sports attitudes scale at baseline. In the case of missing baseline controls we imputed the missing covariate value with the average (mean for continuous controls and median for discrete) of the non-missing observations and this imputation was accounted for with a dummy variable equal to one for imputed observations. We allowed for clustering of the random error term at the cluster level. P-values are two-sided and constructed using a t-bootstrap accounting for stratification by region and treatment status. RW P-values use the Romano-Wolf method to account for multiple hypothesis testing across all outcomes measures in table.

Table A13: Participation in Sports

	Control Mean	Girl Only	AME	P-value	Integrated	AME	P-value	N
Girl participates in sports	0.156	0.621*** (0.131)	0.091	0.000	0.523*** (0.129)	0.074	0.001	5146

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. P-values are two-sided and constructed using a t-bootstrap accounting for stratification by region and treatment status. RW P-values use the Romano-Wolf method to account for multiple hypothesis testing across all outcomes measures in table.

Table A14: Mediation Analysis

	(1) Attending School	(2) Married | 15-17 at BL	(3) Married | 15-17 at BL
Girl only arm treatment	0.419*** (0.154)	-0.254* (0.132)	-0.171 (0.134)
Integrated arm treatment	0.408*** (0.147)	-0.203 (0.143)	-0.140 (0.143)
Attending School			-1.541*** (0.211)
Observations	2715	2715	2715
P-value: diff. in Girl only coef			0.00517
P-value: diff. in Integrated coef			0.0337

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. P-values are two-sided and constructed analytically, accounting for clustering.

APPENDIX B: ADDITIONAL TABLES

Table B.1: Baseline Characteristics and Balance

	Full BL Sample	Analysis Sample	Control (Analysis Sample)	Girls' Only (Analysis Sample)	Integrated (Analysis Sample)
Age	14.358 (1.703)	14.297 (1.692)	14.302 (1.665)	14.287 (1.709) [p=0.527] [$RW\ p$=0.846]	14.303 (1.702) [p=0.824] [$RW\ p$=0.997]
Carer's Years of Education	1.097 (2.545)	1.100 (2.522)	1.314 (2.750)	0.885 (2.353) [p=0.078] [$RW\ p$=0.509]	1.085 (2.414) [p=0.190] [$RW\ p$=0.766]
SC/ST Caste	0.359 (0.480)	0.352 (0.478)	0.322 (0.467)	0.339 (0.474) [p=0.639] [$RW\ p$=0.846]	0.396 (0.489) [p=0.140] [$RW\ p$=0.710]
Asset Index	-0.053 (0.870)	-0.038 (0.874)	0.007 (0.914)	-0.062 (0.855) [p=0.202] [$RW\ p$=0.682]	-0.060 (0.849) [p=0.256] [$RW\ p$=0.840]
Attending School	0.758 (0.428)	0.777 (0.416)	0.806 (0.395)	0.779 (0.415) [p=0.371] [$RW\ p$=0.846]	0.745 (0.436) [p=0.029] [$RW\ p$=0.245]
Marriage Fixed or Engaged	0.019 (0.138)	0.012 (0.110)	0.005 (0.071)	0.011 (0.106) [p=0.110] [$RW\ p$=0.588]	0.021 (0.142) [p=0.001] [$RW\ p$=0.012]
Married but Gauna not Performed	0.013 (0.112)	0.009 (0.092)	0.008 (0.091)	0.010 (0.101) [p=0.113] [$RW\ p$=0.588]	0.007 (0.084) [p=0.645] [$RW\ p$=0.997]
Patriachal Gender Norms Factor	0.003 (0.989)	0.000 (1.000)	-0.026 (0.997)	0.096 (1.030) [p=0.150] [$RW\ p$=0.637]	-0.067 (0.966) [p=0.653] [$RW\ p$=0.997]
Carer's Patriachal Gender Norms Factor	-0.003 (0.986)	0.000 (1.000)	-0.030 (1.001)	0.068 (1.013) [p=0.184] [$RW\ p$=0.682]	-0.035 (0.984) [p=0.982] [$RW\ p$=0.997]
Self Esteem Factor	-0.012 (0.984)	-0.000 (1.000)	0.088 (0.982)	-0.140 (1.024) [p=0.006] [$RW\ p$=0.055]	0.045 (0.981) [p=0.677] [$RW\ p$=0.997]
Self Efficacy Factor	-0.014 (0.999)	-0.000 (1.000)	0.111 (0.985)	-0.074 (0.977) [p=0.034] [$RW\ p$=0.274]	-0.044 (1.028) [p=0.148] [$RW\ p$=0.710]
Mental Health Factor	-0.012 (0.986)	0.000 (1.000)	0.025 (1.000)	-0.037 (0.998) [p=0.432] [$RW\ p$=0.846]	0.010 (1.002) [p=0.883] [$RW\ p$=0.997]

Notes: Table presents means and standard deviations (in parentheses) by treatment group. P-values test the difference between each treatment group mean and the control. P-values are two-sided and constructed using a cluster t-bootstrap accounting for stratification by region and treatment status. RW P-values use the Romano-Wolf method to account for multiple hypothesis testing across all outcomes measures in table.

1

Table B.2: Differential Attrition

Dependent Variable:	Attrition	Attrition
Girl Only	0.0712	0.0560
	(0.117)	(0.118)
Integrated	0.0750	0.0194
	(0.131)	(0.134)
Observations	5927	5921
Controls	no	yes

Notes: Table presents logit coefficients and standard errors (in parentheses, corrected for clustering) for regressions of attrition on treatment status and, in the second column controls for girls age and education (through a full set of interactions between the age in years and a dummy indicating if the girl was in school at baseline), household wealth and caste at baseline. $*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$. P-values are two-sided and constructed using a cluster t-bootstrap accounting for stratification by region and treatment status.

Appendix C: Pre-Analysis Plan

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.

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

8

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

9

	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

11

			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:

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

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:

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

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.

APPENDIX D: SAMPLE SESSION

Session 3: All about Power

Results of study

After this session, participants:

- Will be able to understand power and will know how it affects our life.
- Will be able to understand that how power affects anyone's personality and development.

Notes for the mentor

With Power Walk activity, sometimes young age girls feel that they are being isolated intentionally, therefore trainer should be prepared mentally to end the session with positive feelings.

Time: 90 minutes

Required items: flipchart, few markers

Activity 1: Enacting Power

- Divide the members in three groups. Tell them that each group will play its role. Ask group to play situations given below.
- Give each group one from the situations given below –
    - How is power seen at home among different people?
    - How is power in school seen among different people?
    - How is power seen among different people in the community?

- Provide 5 minutes to each group for preparations and give 2 minutes for presentation. After each presentation, ask questions related to discussion.

Questions for discussion

For the group, which are watching the presentation, after each presentation:

- Who has power in this situation?
- In the role play, people who had power what they did of this power?
- The people who had power in role play, is it normal thing to have power for all these people at your home, school and community? Does it happen everywhere?
- Can the characters shown in role play use their power in different way?

When everyone's presentation ends, then for everyone:

- In all presentations the people who have power, do they have some similarities?
- Is it bad to have power for someone?
- Can power be used for both good and bad works?

Issues for discussion

- Power is not bad in itself. Power means strength and how is strength utilized, it makes a difference. In this activity, teenagers saw that in our life and relations different people have power in different amount. For example, in someone's family Mother is seen in power and in some cases son is seen in power completely. Similarly we have to take care that power is present in relationships and in each relationship we analyze our power and focus on our powers. And use the strengths obtained from that power to get rights in our life and to promote equality.
- Power can be used in any way for example 'power on someone' and 'power with someone':
  1. **Power on someone –** to control any situation negatively, to suppress, to force, discrimination, corruption, bad behavior. This kind of power detract people's rights and dominates over them.
  Who will have power, this depends on surrounding social environment and structure. We usually see that in our surrounding environment sometimes men dominate in power and somewhere women do, somewhere any specific caste or religion.
  2. **Power with someone –** Like using the collective power gaining strength with other people or group so that everyone can get equality based on

constitutional rights. This kind of power depends on harmony, equality and mutual help and develops from talent and knowledge of different people.

Activity 2: Power Walk

- Divide participants in two groups. Ask one group to assume that they are boys and another group to assume they are girls. Now, make them stand in a straight horizontal line, shoulder to shoulder. Explain that you will be saying some sentences, and if they can do those particular things being a boy or a girl, they should move one step ahead. If not, they should move one step backwards.

Statements:

- I can be head of the family?
- I can stay outside home after 7pm without any reason?
- Can you study after standard 7?
- I cook or support your mother in cooking every day/most of the time.
- I can play outside in a public ground with my friends?
- I can spent a night in my friends house anytime?
- I can visit a religious place whenever I want?
- I can marry someone whom I like?
- I can work outside and earn money to look after my parents and others?

Discussion Questions

- What do you see? Who is ahead? Why?
- Who had to stay behind? Why?
- Why do boys get more privileges?
- Do girls have more restrictions? Why?
- What kind of restrictions do girls have, beside the ones that have been mentioned?
- Are there any restrictions for girls in schools?
- Is it right/proper that girls should have restrictions? Why? Why not?
- What we can do to reduce these restrictions?
- What will happen if girls also get privileges?

Discussion Point

- Unfortunately, a girl is always looked upon as someone who will go away from the family and therefore, not contribute to the family income.
- A boy's education is looked upon as an investment for the family's future needs.

- Boys have privileges while girls face greater restrictions. These restrictions are violation of their rights and hinder their development and growth.
- We should all work towards increasing the privileges that girls have and reduce their restrictions. We should think about how each one of us can contribute towards this at our own personal levels.
- Introduce the term patriarchy here and explain how patriarchy sets certain norms of masculinity and femininity.

Note for the mentor: when you are conducting discussion through asking questions, during that time these issues can be discussed –

- Girls are seen with this perspective that they will go their home after marriage and therefore will not help in family income. That's why it is not spent on their education.
- Studies of a boy are seen as an investment, which will fulfill future needs of the family.
- Boy is given many rights, but girls are restricted in many ways. These restrictions are violation of their rights and it will hamper their growth.
  (Connect examples of last activities to this –like boys are allowed to go out in night time but girls are not. The effect of this is that if girl has move in night for some urgent work, she will not have that confidence and will feel feared, due to this girls are kept away from many opportunities like studying in other cities etc.)
- We should work in this regard that girls get equal opportunities and restrictions imposed on them are reduced. As girls have more restrictions, we have to put special efforts to end this imbalance and inequality. We will have to think that all amongst us, anyhow on personal levels, should be responsible for this cause.
- Take help of handout to mention article 15 of the constitution regarding this.

End of session:

(Note for the trainer/mentor: Ask group member to present in brief whatever they have learned in this session. Take help of points given below to present summary of the session).

- Power is not good or bad in itself. How you use your power makes a difference? Power can be used to suppress people who are weaker than us. While if power is used in right way, then it would make weak people stronger.
- As we discussed in the last session, social gender norms determine that how boys and girls should behave. We could be haunted by violence on not doing so.
- As there are different power structures of rights/power in our society for example, girls face more restrictions, some specific caste is given more power and some are given less, difference is also seen between poor and rich, also power difference is

found between sharp minded and those who are weak in studies. There we should try to reduce inequality and gender discriminations between boys and girls and men and women. The effect of gender based discrimination is deepr on those boys and girls of the society who are already suppressed in social power structures (like poor, illiterate).