

Multi-Level Programming Aimed at Gender Norms Transformation to Improve Capabilities of Young Adolescents in Ethiopia: A Cluster Randomized Control Trial

Pre-analysis Plan

Nicola Jones, Sarah Baird, Joan Hamory, and Tassew Woldehanna

October 9, 2020

Abstract

Adolescence is considered a ‘critical period’ of development, and a window for interventions to improve economic and social well-being. However, it is also a time when unequal gender norms become entrenched, often reinforced by peers, families, communities and broader institutional structures. Without change in norms at each of these levels, improved outcomes for adolescent girls and boys are unlikely to be sustained. This study employs a multi-arm cluster randomized trial to evaluate the impacts of multi-level programming aimed at gender norm transformation on young adolescent empowerment in Ethiopia. Study sites are randomly allocated to: (i) curriculum-based programming with girls, (ii) curriculum-based programming with girls, boys, and parents, (iii) the latter with community-level systems strengthening, (iv) the latter with asset transfers to girls, and (v) control sites. The aim of this evaluation is to fill evidence gaps around impacts of multi-level interventions for young adolescents while also unpacking the added value of specific components.

Keywords: Gender, Adolescence, Capabilities, Norms, Empowerment

JEL codes: O15, O12, I25, I15, I32

Study Pre-registration: This study has been registered on the AEA RCT Registry (#AEARCTR-0004024, <https://www.socialscicenter.org/trials/4024>), as well as on ClinicalTrials.gov (#NCT03890237, <https://clinicaltrials.gov/ct2/show/NCT03890237?id=NCT03890237&rank=1>).

Acknowledgements: We would like to thank our colleagues at Pathfinder, Courtney Boudreau and Tricia Petruney, for their helpful comments and inputs into the design of this evaluation, and our data collection partners in Ethiopia, EDRI and Laterite, for their hard work in gathering the data for this evaluation.

Contents

1	Introduction.....	4
2	Proposed Timeline	7
3	Research Design.....	8
3.1	Basic Methodological Framework	8
3.2	Interventions	8
3.2.1	<i>Her Spaces</i>	10
3.2.2	AWH-E Simple	10
3.2.3	<i>Act With Her - Ethiopia</i>	11
3.2.4	AWH-E + Asset Transfers	13
3.3	Sampling and Randomization	13
3.3.1	Site Selection	13
3.3.2	Site Randomization	15
3.3.3	Adolescent Sampling	15
3.3.4	Variations from the Intended Sample	16
3.3.5	Intervention Implementation and Masking	18
3.3.6	Statistical Power.....	18
3.4	Research Questions.....	19
3.5	Outcomes	19
3.5.1	Girls' Education and Learning Outcomes.....	20
3.5.2	Girls' Bodily Integrity Outcomes.....	20
3.5.3	Girls' Physical Health, Nutrition, and SRH Outcomes.....	21
3.5.4	Girls' Psychosocial Wellbeing Outcomes.....	21
3.5.5	Girls' Voice and Agency Outcomes	21
3.5.6	Girls' Economic Empowerment Outcomes	23
3.5.7	Girls' Cross-Cutting Outcomes.....	24
3.5.8	Boys' Outcomes.....	27
3.5.9	Caregivers' Outcomes.....	27
3.6	Data Collection and Processing	29
3.6.1	Survey Data Collection	29
3.6.2	Data Quality Control and Processing.....	30
4	Empirical Analysis.....	31
4.1	Regression Specification.....	31
4.2	Missing Data	32

4.3	Handling of Outliers	33
4.4	Baseline Balance Analysis	33
4.5	Multiple Outcome and Multiple Hypothesis Testing	33
4.6	Heterogeneous Effects	33
5	References	33
6	Appendices	37
6.1	Appendix A: Additional Figures	37
6.2	Appendix B: GAGE Ethiopia Listing Protocol: Rural Sites	39
6.3	Appendix C: Secondary Outcomes for Analysis	41
7	Administrative Information	44
7.1	Funding	44
7.2	Institutional Review Board	44
7.3	Declaration of interest	44
7.4	Acknowledgements	44

1 Introduction

Adolescence is increasingly recognized as a ‘critical period’ for development, akin to the first 1,000 days of life (Bundy et al., 2017). Adolescents experience rapid physical transformations as well as significant changes in brain chemistry. These changes facilitate the complex thinking and emotional skills required for adulthood, but also require support to help young people avoid risky behaviors and make decisions that will promote their current wellbeing and set them on a positive life trajectory (Dahl et al., 2018). This transitional time is particularly fraught for poor adolescent girls in low- and middle-income countries (LMICs). These girls are often kept home from school and face mobility restrictions in their community, rarely have access to formal employment, and are especially vulnerable due to their emergent sexuality. In addition, adolescence is a time when gender norms and expectations—endorsed and reinforced by families, communities, systems, and institutions—are keenly felt and internalized, determining what young people do and expectations around appropriate behavior (Chung and Rimal, 2016).

Adolescent girls in particular face an array of restrictive gender norms, from child marriage to disproportionate care and domestic work burdens, that ultimately lead to economic disadvantage (Duflo, 2012; Jayachandran, 2015). In many contexts, girls are required to leave school and marry early, abandoning not only their educational and occupational plans but also their peer support systems (ICRW, 2016). This is certainly the case in Ethiopia, the context of this study, where school dropout rates remain high, with only about half of young people completing grade 5 and enrolment rates at less than 10% in upper-secondary school (11th and 12th grades) (MOE, 2017). Furthermore, ‘learning levels are poor and appear to have stalled in recent years’ (RISE, 2018), especially for girls. Adolescence is also a time where girls become particularly vulnerable to sexual and gender-based violence (UNICEF, 2014). For example, in Ethiopia, 10 percent of married girls ages 15-19 have experienced sexual violence and among 15-19 year olds, 47% report having undergone female genital mutilation or cutting (CSA and ICF, 2016). Constrained by the intersection of age and gender, girls also have fewer routes to economic empowerment and less mobility, voice and agency within their homes, schools, and communities (Rahman, Kabir, and Shahidullah, 2007; World Bank, 2016).

As a ‘critical period’ of development, adolescence is a key window of opportunity for intervention (Bundy et al., 2017; Sheehan et al., 2017; Steinberg, 2015; UNFPA, 2014). Programs often targeted at adolescent girls include life skills and vocational training, school-based interventions, cash transfer or financial incentive programs, information and awareness campaigns, role models, and health services and counseling (Baird, McIntosh, and Özler, 2019; McQueston, Silverman, and Glassman, 2013). Other interventions targeted at this group are variants of interventions in these broad categories, such as the Empowerment and Livelihood for Adolescents (ELA) program in Uganda which provides life and vocational skills through mentors in adolescent development clubs (Bandiera et al., 2020) or the peer-led

sessions in Safe Spaces in Bangladesh combined with (in-kind) incentives to delay marriage until the legal age of 18 (Buchmann et al., 2018). There are also programs that more broadly target gender norm change at the parent or community level, including TESFA in Ethiopia (Edmeades and Hayes, 2014).

Research in development economics on programs targeted at adolescent girls has shown evidence on improved educational outcomes, especially increasing enrollment and attendance, and to some extent improving test scores. Özler (2016) summarizes evidence on such interventions, including those that (i) increase the returns to schooling, (ii) reduce the direct, indirect, and opportunity costs of schooling and (iii) provide financial support (either through cash or in-kind) to parents and adolescent girls. Similarly, a variety of interventions have been shown to be effective in reducing child marriage and teen pregnancy, and increasing the use of contraceptives and knowledge of sexual and reproductive health. These interventions not only include (conditional and unconditional) cash transfer programs (e.g. Baird, McIntosh, and Özler, 2011), but also a large number of girl-focused programs that provide various combinations of life skills, vocational skills, and educational and financial support to adolescent girls and their families (e.g. Bandiera et al., 2020). The evidence on Technical and Vocational Training Programs (or TVET), typically targeting out of school adolescents, remains mixed (Chakravarty et al., 2015; Blattman and Ralston, 2015), but Chakravarty, Haddock and Botea (2015) conclude that for girls the most promising programs take place in girl only or girl-friendly settings, providing a combination of SRH information and complementary training and assets.

The existing literature on gender norms in development economics has largely centered on adults (Beaman et al., 2009; Jensen and Oster, 2009; La Ferrara et al., 2012). While many of the programs discussed above focus on adolescent girls, gender norms are for the most part not at the center of programming. That said, there is an increased interest in the specific role of gender norms in adolescence (see, for example, Dhar, Jain and Jayachandran, 2018). The present study seeks to grow the evidence base in this area by evaluating the impact on empowerment of adolescent programming that takes a complex, multi-level approach.

The aim of this study, with the research component funded by the UK Department for International Development and the intervention component by the Bill & Melinda Gates Foundation, is to evaluate *Act with Her-Ethiopia* (AWH-E), a two-year multi-level program including curriculum-based programming aimed at gender norm transformation and empowerment for young adolescent girls and boys (aged 11-13), and their primary caregivers, as well as community-level engagement and systems strengthening. Through a multi-arm cluster randomized trial, AWH-E will be compared to a pure control group as well as three other interventions: a pre-existing standalone girls' group program (*Her Spaces*), a more basic gender-synchronized program (AWH-E Simple, without community engagement and systems strengthening), and the multi-level program design with economic support for girls (AWH-E + Asset Transfers). This

evaluation will fill critical evidence gaps around (i) the potential value of beginning interventions with very young adolescents; (ii) the added value of specific components of complex interventions such as asset transfers and community level systems strengthening; and (iii) the short- and medium-term impacts of multi-level interventions that seek to transform gender norms and support adolescent girls' healthy transitions across a range of domains.

In addition, this study will add to the evidence base by moving beyond simply measuring gains in education, health, and/or income-generating activities, and instead attempting a more comprehensive measure of adolescent empowerment embedded in the *Gender and Adolescence: Global Evidence* (GAGE) conceptual framework (GAGE Consortium, 2017) (See Appendix Figure A.1). Gender inequities and power imbalances in families, communities, and institutions drive poor outcomes—especially for girls—across all aspects of physical health, education, bodily integrity, psychosocial well-being, voice and agency, and economic empowerment, the six primary domains within the GAGE framework that will form the basis of our empowerment measure.

Our approach internalizes Kabeer's (1999) definition of empowerment as “the process by which those who have been denied the ability to make strategic life choices acquire such ... ability.” She argues that two elements—resources and agency—determine an individual's ability to exercise choice and realise their full capabilities. In order for programs to improve empowerment they must either improve agency (voice, participation, decision making), increase resources (financial, human, social, psychosocial or physical) available, or redefine norms and institutions (Golla et al. 2011). This framework suggests that a multi-level intervention such as AWH-E is necessary to move the needle on empowerment. Moreover, Corbin and Heckman (2016) observe that life skill formation is inherently dynamic, and thus investments across the life course are going to be critical to sustained effects on empowerment. Building on this concept, Kabeer (2018) notes that “Many of the efforts to build women's capabilities would have been far more effective... if they had been implemented in an earlier stage in their lives.” The idea being that if investments in critical capabilities took place during childhood and adolescence, it would have provided these women with a solid foundation on which to build subsequent skills and exercise greater agency as adults. The increasing recognition that adolescence—as with the first 1000 days—is a critical period of transition suggests that interventions that tackle the multitude of disadvantages that adolescents face may be particularly effective. The evaluation proposed here provides an opportunity to contribute to the knowledge base on the effectiveness of multi-level programming during adolescence.

2 Proposed Timeline

Site selection, treatment randomization, adolescent sample selection, and baseline data collection were conducted during November 2017 – February 2018; more details on these processes are available in Section 3.3. Table 1 indicates the timeline for programming implementation (which began in early 2019) and evaluation follow-up data collection (final column). The first round of follow-up data collection is currently ongoing, and we expect to launch the second follow-up round in mid-November 2020.

Table 1. Timeline of Programming and Evaluation

		AWH/HS recruitment	AWH/HS adoles. groups	AWH asset transfers	AWH parent groups	AWH systems strengthening	GAGE data collection
2019	Jan						
	Feb						
	Mar						
	Apr						
	May						
	Jun						
	Jul						
	Aug						
	Sep						
	Oct						
	Nov						
	Dec						
2020	Jan						Follow-up #1
	Feb						
	Mar						
	Apr						
	May						
	Jun						
	Jul						
	Aug						
	Sep						
	Oct						
	Nov						
	Dec						
2021	Jan						Follow-up #2
	Feb						

3 Research Design

3.1 Basic Methodological Framework

This study is a parallel 5-armed cluster randomized control trial that seeks to compare key outcomes across the six GAGE capability domains (education, bodily integrity, health, psychosocial well-being, voice and agency, and economic empowerment; see Appendix Figure A.1) for approximately 4,000 girls and boys aged 11-13 years (at the start of programming in early 2019) in rural Amhara and Oromia regional states, as well as caregiver behaviors, attitudes, and aspirations toward these adolescents. Impacts will be evaluated after approximately 8 months of programming (late 2019) and approximately 20 months of programming (late 2020) between communities (kebeles) randomly assigned across the following control and treatment arms: (a) pure control, (b) *Her Spaces*, (c) *Act with Her-Ethiopia* (AWH-E), (d) AWH-E and Asset Transfers, and (e) AWH-E Simple. A cluster randomized control trial was deemed appropriate for this evaluation as there is sufficient uncertainty about the potential impact of the various interventions, in particular in comparison to each other. In addition, given the community level nature of some intervention components, clustering was deemed essential. The specific set of intervention arms was designed through close collaboration among researchers in the GAGE consortium, the project consortium (led by Pathfinder International and CARE International, with input from Save the Children US), and the Ethiopian government. Regression analysis will be used. In what follows, the research design is described in detail.

3.2 Interventions

Before detailing the set of interventions under study, we further rationalize the need for the multi-level programming evaluated. Unequal gender norms and power dynamics are often reinforced by adolescents' peers, families, communities and the broader institutional structures that surround them. Thus, without change in gender attitudes and norms at each of these levels, improved adolescent outcomes are much less likely to be sustained.

There is a growing literature that highlights the importance of working with adolescent girls to tackle the deep-rooted effects of discriminatory gender norms and to support their individual and collective empowerment (Harper et al., 2018). Working with adolescent boys, before gendered attitudes and behaviors are firmly cemented, is also likely to be critical. Because gender is a relational concept, this work is important not only in terms of the space it opens for girls, but also in terms of impacts on boys' own lives (Kato-Wallace et al., 2016). Rigid gender norms place pressure on boys to prove their masculinity, and can drive them to engage in harmful behaviors such as violence towards girls and other boys, unsafe sex, and substance abuse. Indeed, interpersonal violence is a leading cause of mortality for adolescent boys globally, second only to road injuries (WHO, 2014).

Families and communities also often perpetuate inequitable gender norms. Parents, siblings, and broader communities both directly and tacitly communicate collectively-held expectations and beliefs about how girls and boys, and women and men, should behave and interact in specific social settings at specific stages of their lives. Parents and other adult family members also often make decisions that affect girls' healthy transitions (e.g., school dropout, child marriage) and impact girls' mobility. Through social pressures and standards communities either inhibit or foster progress toward gender equality.

Finally, institutional structures can further drive gender inequality and reinforce other structural drivers of girls' constrained choices and unhealthy outcomes as they transition through adolescence. Formal laws and policies, such as those that allow or prevent child marriage or deny girls' inheritance rights often uphold and promote unequal treatment of women and girls. Furthermore, services that are critical for girls to positively transition through adolescence, such as education, health and psychosocial services, and SGBV prevention and response, are often limited in coverage, of poor quality, do not address gender equity, and are not responsive to the needs of adolescents. To have the most sustained impact, institutional structures (including education, health, justice, and social protection) must be well coordinated and work cohesively to support girls and women. When this is not the case, advances made by one system are less effective than they could be if supported by other institutional structures.

In order to contrast the impacts of programming targeted at these different levels, we employ a multi-arm parallel cluster randomized control trial to evaluate the impacts of different intensities of programming aimed at gender norm transformation on young adolescent outcomes in Ethiopia across six different capability domains, as well as on caregiver behaviors, attitudes, and aspirations toward these adolescents. Study sites are randomly allocated to one of five different study arms: (i) pure control, (ii) curriculum-based programming to girls only, (iii) curriculum-based programming to girls, boys, and parents, (iv) curriculum-based programming for girls, boys, and parents as well as community-level systems strengthening, and (v) curriculum-based programming for girls, boys, and parents, with community-level systems strengthening as well as asset transfers for girls.

Table 2 provides a brief description of each non-overlapping intervention arm evaluated in this cluster RCT. These interventions are then described in greater detail below the table. Note that there is also a control study arm.

Table 2: Intervention arms¹

Arm
1: <i>Her Spaces (HS)</i>: Curriculum-based programming for 11-13 year old girls for 10 months; standard parental and community engagement for 10 months.
2. <i>Act with Her – Ethiopia curriculum only (AWH-E Simple)</i>: Curriculum-based programming for 11-13 year old girls and boys for 10 months; 6 dedicated sessions with parents over 8 months.
3: <i>Act with Her – Ethiopia (AWH-E)</i>: Curriculum-based programming for 11-13 year old girls and boys for 10 months; 6 dedicated sessions with parents over 8 months; and community-level system strengthening up to 24 months.
4: <i>Act with Her - Ethiopia with asset transfer (AWH-E + Asset Transfers)</i>: Curriculum-based programming for 11-13 year old girls and boys for 10 months; asset transfers for girls over 5 months (total value of USD 115); 6 dedicated sessions with parents over 8 months; and community-level system strengthening up to 24 months.

3.2.1 *Her Spaces*

Her Spaces comprises a set of activities leveraging the curriculum and program model originally funded by the UK Department for International Development and implemented by Girl Effect, IPE Global, and Pathfinder International. *Her Spaces* is a girl-focused, safe spaces program that works with girls to provide them with the skills and supportive network to successfully navigate adolescence. Girls between the ages of 11 and 13 participate in 10 months of weekly curriculum-based, near-peer mentor-led group sessions (40 sessions). The curriculum covers the domains of health, nutrition, education, safety, economic security, and voice and participation.

In the *Her Spaces* intervention arm, there is a fairly standard light-touch engagement strategy with parents and male peers of program participants, as well as community stakeholders. Sensitization meetings to raise awareness of and familiarity with the program are held at the kebele level, approximately twice during the programming implementation phase, to discuss the intent of the program and the main themes covered in the girl group meetings.

3.2.2 *AWH-E Simple*

This arm (*AWH-E Simple*) implements the *Act With Her* curricula described below with 11-13 year old girls and boys over 10 months. The adolescent groups follow the same model (near-peer, same-sex mentors, 40 weekly meetings for girls, 18 meetings for boys, 4 of which are joint sessions for boys and

¹ Note that *Her Spaces* and *Act With Her* programming includes all adolescents of the relevant gender aged 10-13 in program sites. The GAGE evaluation will only focus on adolescents aged 11-13 at the time of program launch, however, as this was the group for which baseline data was collected. Throughout this document we refer to programming as including those aged 11-13 years to avoid confusion.

girls, 6 sessions for parents), but there are no community-level systems strengthening activities (e.g., SAA groups or the CSC approach) or asset transfers.

3.2.3 Act With Her - Ethiopia

Act With Her – Ethiopia (AWH-E) is designed to build from the foundations of the *Her Spaces* model to facilitate adolescent transitions to adulthood via improvements across the six GAGE domains (education, bodily integrity, health, psychosocial well-being, voice and agency, and economic empowerment). Project activities fall into three main categories: curriculum-based programming with adolescent girls and boys, community transformation, and systems strengthening. In particular, AWH-E includes group sessions with 11-13 year old girls and boys for 10 months, and community-level and high-level system strengthening work for up to 24 months.

AWH-E has developed project-specific curricula aimed at gender norm transformation for boys and girls, which comprehensively address key topics across the six GAGE domains.² The curricula have been refined and adapted to the Ethiopian context. Girls of eligible age participate in 10 months of weekly curriculum-based, near-peer mentor-led group sessions (40 sessions). Near-peer male mentor-led groups for boys of the same eligible ages run concurrently, meeting approximately twice a month (18 sessions). Topics covered in boys' and girls' groups are temporally aligned, and four sessions bring the boys' and girls' groups together.

Additionally, the AWH consortium has developed a parents' guide which outlines six sessions to orient parents (or any primary caregiver for the AWH adolescents) to the topics covered in the adolescent curricula and help them to create a supportive environment for their adolescent. Parent sessions are not segregated by the gender of the parent, but sessions are held separately for the parents of boy adolescents and parents of girl adolescents. The approach is designed to encourage both male and female parents/caregivers to participate together, as the program team feels that it is important for fathers to engage jointly with mothers in issues related to their daughter, and vice versa for mothers and sons.

Applying a socioecological approach which recognizes the influence that social norms and systems have on adolescent health, AWH-E also includes high-level systems strengthening activities. This entails strategically engaging key stakeholders across multiple sectors at the national, regional, and woreda levels³, with the objective of raising the visibility, prioritization, and subsequent improvement of adolescent-responsive systems and services. A secondary objective is to improve the levels and frequency of dialogue and connections between stakeholders in multiple sector systems, ideally leading to improved coordination

² These curricula will be publicly available after the AWH programming has completed. More information on the AWH programming can be found at <https://www.pathfinder.org/publications/act-with-her-project-factsheet/>.

³ To the extent that systems strengthening work is at the woreda level or higher, this will also impact the control and *Her Spaces* kebeles. Outside of this systems strengthening work, there is no implementation in the control kebeles.

and linkages between them in ways that better serve the needs of adolescents. Policymakers and other key decision-makers who are potential future AWH implementers and eventual end-users of the research (including representatives from youth-led organizations) help generate political will and also advise the program during key reflection points during implementation. Through collaboration with these multi-sector actors, AWH seeks out windows of opportunity for catalyzing enhanced adolescent-related policies, guidelines, or capacity for service delivery. Key systems targeted for this high-level stakeholder engagement include health, education, sexual- and gender-based violence (S/GBV), and child protection. The proposed focus of national-level systems strengthening includes policy dialogues, stakeholder engagement, and multi-sector networking regarding child protection, justice, education (violence reduction), and health (menstrual hygiene management).

The community-level systems strengthening activities focus on activities that a) establish locally-led mechanisms for discussing social norms in ways that initiate shifts in them over time (including past the project timeline), and b) enhance local capacity for social accountability. Work on catalyzing social norms shifts is primarily focused on applying CARE's *Social Analysis and Action* (SAA) approach to gender and social transformation.⁴ Implemented for more than 10 years in multiple program models by CARE, SAA seeks to enable communities to identify for themselves the linkages between social factors and well-being, and then determine actions that will help improve them. The approach allows communities to lead reflection on gender, social, and power norms through participatory tools and discussion. The AWH SAA group meet monthly to discuss harmful socio-cultural norms relevant to their local community, identified and agreed upon beforehand by a diverse set of key 'power holders' and gatekeepers in the community. After discussing the norms and reflecting on what elements of health or well-being they may be negatively contributing to, the group devises an action plan as to how they can be combated. The SAA groups are launched approximately 2 months after the first girls' groups start and will run for up to 2 years in each kebele.

The objective of using SAA in *Act With Her* is to spark shifts in the enabling environment that will complement and positively reinforce the improved agency, knowledge, and behaviors demonstrated by adolescents. The proposed focus of community-level systems (and services) strengthening is as follows: enhance menstrual hygiene management and violence reduction within schools; improve local child protection and anti-harmful traditional practice measures (e.g., raising visibility of violence response mechanisms and/or providing age/gender sensitivity training to local officials); foster better linkages between school, health, and social protection platforms; enhance social systems and norms via SAA and community score card activities (the latter for health clinics and schools).

⁴ More details on the SAA approach can be found at https://insights.careinternational.org.uk/images/in-practice/Gender-in-the-workplace/SAA.GlobalImplementationManual_FINAL.English.rights-reserved_2018.pdf.

Strengthening local capacity for social accountability is approached through the implementation of CARE's *Community Score Card* (CSC) intervention.⁵ The main goal of the CSC is to positively influence the quality, efficiency and accountability with which services are provided in a particular community. Used throughout CARE's programming, the CSC offers a way to increase participation, accountability and transparency between service users, providers and decision makers. Moreover, the government of Ethiopia has its own CSC used to solicit feedback on health services. In AWH, a CSC will be used that blends the format of the government and CARE templates (focused on health service delivery) to also include considerations for education, S/GBV, and/or child protection services. Particular attention is paid to ensuring that adolescent girls and boys directly participate in the CSC processes. The objective of CSC in AWH is to improve local stakeholders' ability, including young people themselves, to hold providers of key services for adolescents accountable for optimal access and quality.

3.2.4 AWH-E + Asset Transfers

This arm (AWH-E + Asset Transfers) implements the AWH-E program described above, using program curricula with 11-13-year-old girls and boys over 10 months and conducting community-level systems strengthening for up to 24 months. Additionally, the girls participating in this arm receive one of three asset transfer options. The packages are of equal value (115 USD) – one includes school supplies, one includes hygiene supplies, and the last is a combination of the first two. Each adolescent girl chooses which package she would like to receive and receives components of the package at 3 timepoints over the course of the first 10 months of the project, as indicated on the timeline in Table 1.

The asset package is introduced at recruitment, and the adolescents select their package within the first weeks of the adolescent group meetings. Adolescents who do not show up to these meetings are assigned the combined package. The first package is transferred near the start of the school year (late August), and the second and third spread in the second half of the program (see Table 1). The timing of only the first transfer is announced. If a girl misses a transfer day, she only receives missed transfers at the conclusion of the program if she attended at least three-quarters of the program meetings.

3.3 Sampling and Randomization

3.3.1 Site Selection

This study is operating in the Amhara and Oromia regions of Ethiopia, areas that are both heavily populated and where child marriage is common for girls. Oromia is the most populous region in Ethiopia with a

⁵ More details on the CSC approach can be found at https://www.care.org/sites/default/files/documents/FP-2013-CARE_CommunityScoreCardToolkit.pdf.

population of approximately 38.3 million and, according to the Ethiopian DHS 2016, a median age of first marriage of 17.4 years (among women aged 20-24 years). Amhara is the second most populous region in Ethiopia with a population of approximately 28.8 million (assuming 2018 national population of 107 million), and according to the Ethiopian DHS 2016 a median age of first marriage of 16.2 years (among women aged 20-24 years).

Rural sites were selected within each of these regions on the basis of two key but complementary considerations: (1) a review of existing data and evidence on adolescents and gender in Ethiopia, which highlighted the importance of understanding both the economic and social drivers that underpin disadvantage (Stavropoulou and Gupta-Archer, 2017a; Jones et al., 2017; Stavropoulou and Gupta-Archer, 2017b); and (2) programming capacity (including operational presence and experience, and an absence of security concerns) on the part of the NGO implementing partners, Pathfinder (in Amhara) and Care Ethiopia (in Oromia). In terms of the first consideration around vulnerability criteria, we selected geographical areas with high rates of child marriage (as a proxy for conservative gender norms) and those with the greatest proportion of ‘hotspot’ child marriage districts (see Jones et al., 2016), as well as those that were economically disadvantaged and/or food insecure.

In particular, two zones were selected for study inclusion (one from each region) on the basis of food insecurity status and high rates of child marriage (especially among the 10-14 age bracket), as a proxy of conservative gender norms and relatively less programming on child marriage. South Gondar Zone (Amhara) has one of the highest rates of child marriage in the region, and has had relatively less child marriage programming compared to other zones in the region (e.g. compared to Gojjam, which was the focus of both Population Council’s Berhane Hewan program and DFID and the Ministry of Women, Children and Youth (MOWCY)’s Finote Hiwote End Child Marriage program). East Hararghe Zone (Oromia) was selected because a number of districts in the zone were identified as ‘hotspot’ *woredas* with exceptionally high child marriage rates, and the zone itself had among the highest rates of child marriage for 10-14 year olds and 15-17 year olds in Ethiopia, as identified in the Government of Ethiopia Alliance to End Child Marriage Assessment carried out by ODI in conjunction with MOWCY and UNICEF in 2015. Appendix Figure A.2 provides a map of Ethiopia, with the two zones in the study sample highlighted (in dark purple).

Five *woredas* (districts) were purposely selected within each zone on this same basis, alongside operational presence of the planned programming implementors; these include Ebenat, Lay Gayint, Libo Kemkem, Simada, and Tach Gayint in South Gondar Zone, and Fadis, Babile, Jarso, Haramaya and Gursum in East Hararghe Zone. Within these 10 *woredas*, all *kebeles* (communities) were characterized into three groups: (i) unsafe for data collection and programming, (ii) marginalized (lack of programming, isolated from key services and road/transport infrastructure) and (iii) less-marginalized (in terms of access to

services and to the main woreda town). Kebeles identified by local officials as characterized by high security concerns were excluded from consideration for study inclusion. Among the remaining eligible kebeles, sixteen kebeles (6 marginalized, 10 less-marginalized) in each woreda were randomly selected (using Microsoft Excel) to be included in the study.⁶ The study includes a total of 155 kebeles (58 marginalized, 97 less-marginalized) in ten woredas.

3.3.2 Site Randomization

Prior to the start of baseline data collection, the 155 kebeles were block randomized (by woreda, and kebele marginalization status) into control and four hypothetical treatment arms (the specific arms were not known at the time of this initial randomization). Randomization was done by creating a random number in Microsoft Excel and then ordering by woreda, marginalization status and then the random number. It was later decided, prior to any analysis within arm, that within each of these blocks, the first kebele listed was the control, and then the four treatment arms after that (*Her Spaces, Act With Her-Ethiopia* (AWH-E), AWH-E Simple, and AWH-E + Asset Transfers). Randomization within each woreda and by kebele marginalization status ensures balance on these two critical observables. In almost all kebeles there was one extra kebele that was assigned to the control group.⁷ At the time of baseline data collection, due to security issues in fourteen kebeles that made data collection impossible, replacement kebeles were used and took on the treatment status of the replaced kebele.

3.3.3 Adolescent Sampling

Although the randomization is performed at the cluster (kebele) level, the unit of analysis in this study is the adolescent. Adolescents were identified through a household census (listing) that was performed in each of the 155 kebeles selected for inclusion in the study immediately prior to baseline data collection during November 2017 through February 2018. Because kebeles are quite large, a pre-defined protocol was followed by the listing teams to select one or more smaller administrative areas within a kebele that would be fully listed, rather than listing entire kebeles. This protocol is included in Appendix B. To identify the sample of adolescents, a household census (listing) was carried out in each selected got/sub-kebele (sub-areas of the larger kebele) according to this protocol. Enumerators went door-to-door, collecting basic contact information for the household as well as a complete listing of all adolescents aged 10–12 living there. Once the listing was complete, the population of households for consideration in the study was

⁶ In two woredas there were only 5 marginalized kebeles, so 11 less-marginalized kebeles were selected, and in one location there were only a total of 11 eligible kebeles (6 marginalized, 5 less-marginalized).

⁷ The rationale for assigning extra kebeles to the control group is that GAGE also has observational data collection ongoing in urban Ethiopia and wanted a larger control group for relevant panel analysis.

restricted to only those households that had at least one eligible adolescent. If the household had more than one eligible adolescent, one adolescent was randomly selected to be the designated eligible. Thus, the evaluation includes only one adolescent per household.

A total of 15 girls and 11 boys were randomly sampled from this census list in each kebele. After accounting for (randomly chosen) replacements, 4,000 individuals were included in the final sample. A total of 3,962 individuals (2,281 girls and 1,681 boys) completed a baseline interview (98% of the originally selected sample size). Replacements households were used in the following cases: (a) refusal on the part of the adolescent to participate, (b) refusal by a parent/guardian to allow the adolescent to participate, (c) adolescent was found to be out of the 10-12 year old age range at the recruitment visit, (d) adolescent was found to be a gender other than what was noted during the census exercise, (e) adolescent was found to be deceased or severely disabled/ill (in a way that would not permit them to participate in the survey) at the recruitment visit, and (f) adolescent had migrated from the area at the time of the recruitment visit (and was not expected to return during data collection period). The refusal rate was low, at approximately 2% of the original random sample.

3.3.4 Variations from the Intended Sample

There are two key issues that have the potential to threaten analysis from the outset: non-random attrition, and non-compliance with the assigned treatment. We address each of these in turn here.

This study focuses on adolescents aged 10-12 at the time of baseline data collection in late 2017 - early 2018, who were then aged 11-13 at the time of programming launch. The research team conducted an interim data collection during the second half of 2018 (prior to the start of programming) in order to assess mobility rates in a subset of sample kebeles, and migration rates (where the core adolescent respondent, or the household as a unit, could not be interviewed at their baseline residential location) were nearly zero.

That said, we did expect migration to increase slightly by the first follow-up data collection round launched in late 2019, given the likelihood that the older of the adolescents will have moved on to lower secondary school, or potentially will have left school to work or to marry. Members of the study team (Baird and Hamory) have worked on other large scale panel data collection efforts in sub-Saharan Africa that involve tracking adolescents and young adults over multiple rounds of data collection. In particular, Baird and Hamory worked on the Kenya Life Panel Survey (Baird, Hamory, and Miguel, 2008; Baird et al., 2016), and Baird on the Malawi Schooling, Income, and Health Risk project (Baird, McIntosh, and Özler, 2019). We are confident that we have put into place protocols that will lower attrition, nevertheless the risk of differential attrition across treatment arms is a possibility. To this end, we will conduct attrition analysis in both the first and second follow-up data collections, looking at overall differential attrition by treatment arm, as well as interactions of treatment with a pre-specified set of baseline characteristics, including age,

indicator for female adolescent, indicator for multiple eligible adolescents per household, household size, indicator for household head literate, indicator for household is female headed (if included as a control – see footnote below), household asset index, and an indicator for whether the household received benefits from the large scale social safety net program in Ethiopia known as the Productive Safety Net Programme (PSDP). If differential attrition by treatment arm is problematic (either in levels or by baseline characteristic), we will follow Karlan and Valdivia (2011), using three different methods to address the potential for differential attrition: Lee (2009) bounds, Kling and Liebman (2004) bounds and Inverse Probability Weighting (IPW). While the Lee bounds rely on an assumption of monotonicity, these two additional methods do not.

Another issue of concern is non-compliance with treatment group assignment. Programming recruitment was done separately in early 2019 by Pathfinder and Care program implementation staff in intervention sites, in a similar manner to study recruitment (using a household listing methodology). Program staff sought all adolescent girls and boys (where appropriate for the intervention arm) in households aged 10-13. There were no other inclusion or exclusion criteria. While there will be a small subset of adolescents interviewed in the GAGE baseline data collection who have since aged out of the programming or moved out of the local area, the programming implementers and GAGE researchers have worked closely together to ensure that the vast majority of GAGE baseline respondents (over 95%) are eligible for programming. This has also since been confirmed with programming registration data.

Pathfinder and Care trained program staff on the importance of compliance with kebele assignment. Mentors were only trained on the curriculum for the kebele they were assigned to and were not told in detail about the activities in other arms. At the end of the *Act With Her* training, mentors were separated by arm so that arm-specific information could be provided only to the mentors working in that arm. To prevent potential confusion among mentor supervisors, Care assigned each mentor supervisor to support kebeles assigned to the same research arm, with Project Officers (who supervise mentor supervisors), assigned to support Arm 1 (*Her Spaces*). Due to logistical concerns in Pathfinder sites (distance between kebeles assigned to the same arm), it was not possible to assign mentor supervisors to support kebeles assigned to only one intervention arm. Therefore, Pathfinder assigned mentor supervisors to support kebeles across multiple arms but provided significant guidance to supervisors to ensure that they understood the importance of maintaining the fidelity of each research arm. When monthly meetings were held among mentors, only mentors from the same intervention arm were brought together.

To reduce the possibility of crossover from the intervention group to the control group, study participants will be assigned to trial groups in the analysis on the basis of the random assignment of the kebele that they lived in at the time of listing.

3.3.5 Intervention Implementation and Masking

The interventions studied in this cluster RCT are being implemented by the international non-governmental organizations Pathfinder in Amhara Region, and Care Ethiopia in Oromia Region. Implementation began in February 2019 with the recruitment of adolescents, and the recruitment and training of group mentors. Timing of all program elements is outlined in Table 1, above.

Study participants were not masked to their treatment group assignment, but were not explicitly told the full set of treatment and comparison groups, or what those other groups would receive. Given the cluster level randomization, contamination should be minimized. Enumerators who collect the follow-up survey data that forms the basis of the evaluation are also not explicitly told the treatment status of a given kebele, however, given the community level nature of the intervention it may be possible for them to ascertain whether it is a treatment or control kebele.

Statistical analyses will be performed by the investigators who will not be masked to the treatment status of the participants. The analysis will be done according to the registered pre-analysis plan to ensure fidelity of research. The trial has been registered at clinicaltrials.gov and the AEA registry.

3.3.6 Statistical Power

This study includes 2,303 adolescent girls and 1,697 adolescent boys in Amhara and Oromia, as well as their primary female caregivers. All individuals are located in 155 kebeles in 10 woredas/districts in Amhara and Oromia, Ethiopia.

We focus power calculations on girls only, and on pair-wise comparisons between one treatment arm and control or between two treatment arms, following the manner in which the research questions are posed in Section 3.4 below. Any comparison that combines one or more arms will have increased power, as will comparisons that include both boys and girls. Assuming power of 80%, 29 kebeles (communities) per-arm, 15 girls per community, an alpha level of 0.05, an intra-cluster correlation of 0.05, and a coefficient of variation of cluster size of 0.5, our minimum detectable effect size is 0.27 – a small to medium effect size.⁸ For boys only, where we have 11 boys per community, our minimum detectable effect size is 0.29, also considered a small to medium effect size. We also perform power calculations for two binary outcomes of interest – experienced peer violence (40% among females at baseline) and enrolled in school (83% at baseline) – and once again assume power of 80%, 29 kebeles per-arm, 15 girls per community, an alpha level of 0.05, an intra-cluster correlation of 0.05, and a coefficient of variation of cluster size of 0.5. Assuming no change in the control group, we are powered to detect a 12 pp decline in peer violence, and a

⁸ CV and ICC come from the baseline data, as do the clusters and number of girls per cluster.

9 pp increase in school enrollment. These effect sizes are reasonable compared to what has been found elsewhere (Bandiera et al., 2020; Buchmann et al, 2018; Dhar, Jain, and Jayachandran, 2018).

3.4 Research Questions

This study seeks to answer four research questions:

- (1) What are the impacts at 8- and 20-months post-program start of a multi-level program aimed at gender norm transformation (AWH-E and component variations) on younger adolescent capability achievements and transitions?
- (2) How do the 8- and 20-months impacts of the multi-level, gender synchronized program (AWH-E and component variations) compare with a program aimed only at adolescent girls (*Her Spaces*)?
- (3) Does the inclusion of community norms transformation and systems strengthening activities enhance the 8- and 20- month impacts of gender-synchronized programming for younger adolescents?
- (4) Does the inclusion of economic support in the form of asset transfers enhance the 8- and 20-month impacts of multi-level gender programming for younger adolescents?

The precisely defined hypotheses we will study are defined in Section 4.1 below.

3.5 Outcomes

Primary and secondary outcomes at 8- and 20-months post-program start (24 and 36 months post-baseline) will be measured for adolescent girls across the six GAGE capability domains and a set of cross-cutting themes (including gender, support, and knowledge). In addition, a number of outcomes across these same domains and themes will be measured at 8- and 20-months post-program start for adolescent boys, as well as for female and (a subset of) male caregivers. Primary analysis for the caregivers is focused on the female caregivers (of both adolescent boys and girls), given sample size considerations.⁹

In what follows, we present primary outcomes by GAGE capability area and cross-cutting theme as measured in the quantitative survey. Complementary qualitative data is also being collected, with the aim of triangulating and nuancing responses (see Jones et al., 2018). It is important to note that the measures included were informed by formative qualitative research undertaken in 2017 in order to strengthen context-relevant questions. Primary outcomes are indicated below in ***bold italics*** (and numbered). We will consider components of the primary outcome indices as secondary outcomes, and may do exploratory analysis on additional secondary outcomes not presented here. A full set of secondary outcomes are presented in Appendix C. Question numbers are indicated for outcomes collected in the Round 2 survey (*in italics*).

⁹ The 8-month follow-up will not collect data from male caregivers.

Note that to construct indices, we employ the following procedure: a) for each index component, create a normalized measure by subtracting the control group mean and then dividing by the control group standard deviation; b) construct the index by calculating the raw mean across all normalized component variables, and c) create the normalized index value by subtracting the control group mean of the index and then dividing by the control group standard deviation of the index. We have signed outcomes such that higher values of indices indicate ‘better’ outcomes. * indicates items that will be measured at 20-months post-program start only.

3.5.1 Girls’ Education and Learning Outcomes

1. ***Index of education***^{*10}: Index consists of three components:
 - a. Indicator for enrolled in school at time in survey (or if school not in session at survey, at time of most recent session) (*CR Module, Education, Q5*)
 - b. Share of school days attended in past two weeks (zero for those not enrolled, missing if no school) (*CR Module, Education, Q5 and Q18-18a*)
 - c. Indicator for did not miss more than one week of school at one time in past 12 months (zero for those not enrolled) (*CR Module, Education, Q5 and Q19*)
2. ***Indicator for aspires to attain a secondary school degree or higher*** (*CR Module, Education, Q20a*)

3.5.2 Girls’ Bodily Integrity Outcomes

3. ***Index of violence***: Index consists of four components:
 - a. Peer violence scale: ranges from 0-6; constructed as the sum of indicators for has experienced –at least once in past 12 months– either in person or digitally, violence from peers, including: (i) hurtful words, (ii) exclusion, (iii) theft/damage of personal property, (iv) physical violence, (v) being forced to do something, or (vi) threats to self or someone close to self; inverse so higher values indicate better outcomes (*CR Module, Violence, Q10a-f*)
 - b. Indicator for did not experience violence in the household, or witness violence against female caregiver, in past 12 months (*CR Module, Violence, Q13-17*)
 - c. Indicator for did not experience sexual violence in past 12 months (*CR Module, Violence, Q24-24a and Q25-25a*)

¹⁰ This index was originally listed as an 8-month outcome as well, but we think it is unlikely to see school enrollment and attendance impacts in the first several months of the programming. We have thus moved this to be a 20-month outcome only.

4. ***Ideal age of marriage***: Measured among those who are still unmarried; code “don’t know” values to median age, code “never” values to max age (*CR Module, Marriage and Relationships, Q40*)

3.5.3 Girls’ Physical Health, Nutrition, and SRH Outcomes

5. ***Index of physical health and nutrition***: Index consists of three components:
 - a. Indicator for self-reported health very good or good (*CR Module, Health and Nutrition, Q1*)
 - b. Proportion of meals yesterday with meat/chicken/fish/egg (zero if no meals yesterday) (*CR Module, Health and Nutrition, Q15-15a*)
 - c. Indicator for has not ever been hungry because not enough food in past four weeks (*CR Module, Health and Nutrition, Q14*)
6. ***Index of SRH practices***: Measured among those who have reached menarche. Index consists of two components:
 - a. Indicator for self-reports normal activities not affected during menstruation (*CR Module, Sexual and Reproductive Health, Q7 and Q10*)
 - b. Index of improved menstrual hygiene practices: ranges from 0-2; constructed as the sum of indicators for (i) has improved menstrual hygiene management (uses re-usable pad, sanitary pad, or other modern sanitary product) and (ii) practices appropriate disposal at home (latrine, rubbish heap, burn, etc) (*CR Module, Sexual and Reproductive Health, Q7 and Q9-9a*)
7. ***Ideal age of first child****: Measured among those who have not yet had a child; code “don’t know” values to median age, code “never” values to max age.

3.5.4 Girls’ Psychosocial Wellbeing Outcomes

8. ***Rosenberg Self Esteem Scale score****
9. ***Patient Health Questionnaire-9 (PHQ-9) score*** (measure of psychosocial distress; inverse) (*CR Module, Social Inclusion, Q19-27*)
10. ***Child and Youth Resilience Scale score*** (*CR Module, Psychosocial and Mental Health, Q13-24*)

3.5.5 Girls’ Voice and Agency Outcomes

11. ***Index of voice and agency***: Index consists of five components:
 - a. Index of participation in decision making: index consists of two components (overall index standardized separately for those in and out of school):

- i. Indicator for has a leadership role in school (among enrolled) (*CR Module, Education, Q5 and Q16*)
 - ii. Index of say in household (0-8, indicators for has a little bit or a great deal of say in how much time spent helping around the house, how much education will get, when to marry, who to marry, who to be friends with, what to do in free time, whether to leave the house, whether to be involved in income generating activities) (*CR Module, Mobility, Voice, and Agency, Q8a-8h*)
- b. Index of comfort expressing oneself: index consists of four components: (standardized separately for those with both parents in home, neither parent in the home, only mother in the home and only father in the home)
 - i. Indicator for comfortable expressing opinion with agemates (*CR Module, Mobility, Voice, and Agency, Q5*)
 - ii. Indicator for comfortable expressing opinion with those who are older (*CR Module, Mobility, Voice, and Agency, Q6*)
 - iii. Index of has discussed issues with mother (0-7, includes indicators for has discussed education, future work, romantic relationships, menstruation, puberty, age of marriage, bullying/harassment at school; missing if no mother in household) (*CR Module, Social Inclusion, Q6a-g*)
 - iv. Index of has discussed issues with father (0-6, includes indicators for has discussed education, future work, romantic relationships, puberty, age of marriage, bullying/harassment at school; missing if no father in household) (*CR Module, Social Inclusion, Q6a-c and Q6e-g*)
- c. Index of voice: standardized separately among enrolled and not enrolled; composed of seven measures, including:
 - i. “My parents or guardians ask for my opinion on things” (0-2, more is better) (*CR Module, Mobility, Voice, and Agency, Q7a*)
 - ii. “My parents or guardians listen when I share my opinion” (0-2, more is better) (*CR Module, Mobility, Voice, and Agency, Q7b*)
 - iii. “My friends ask my advice when they have a problem” (0-2, more is better) (*CR Module, Mobility, Voice, and Agency, Q7c*)
 - iv. “If I see something wrong in school or the neighborhood, I feel I can tell someone and they will listen” (0-2, more is better) (*CR Module, Mobility, Voice, and Agency, Q7d*)

- v. “I can speak up in class when I have a comment or question” (among enrolled) (0-2, more is better) (*CR Module, Education, Q5 and Q15*)
- vi. “I can speak up when I see someone else being hurt” (0-2, more is better) (*CR Module, Mobility, Voice, and Agency, Q7e*)
- vii. “In can ask adults for help when I need it” (0-2, more is better) (*CR Module, Mobility, Voice, and Agency, Q7f*)
- d. Index of mobility: index consisting of five components:
 - i. Indicator for has traveled outside kebele in last 3 months (*CR Module, Mobility, Voice, and Agency, Q1*)
 - ii. Index of not needing permission to go places (0-4, composed of indicators for 4 different locations where permission may be needed--market, homes of relatives/friends/neighbors, religious meeting places, public spaces with friend) (*CR Module, Mobility, Voice, and Agency, Q3*)
 - iii. Index of going different places (0-4, composed of indicators for has gone to any of same 4 different places in the last 3 months) (*CR Module, Mobility, Voice, and Agency, Q2*)
 - iv. Indicator for feels safe walking in community during day (*CR Module, Violence, Q22j*)
 - v. Indicator for feels safe walking in community during night (*CR Module, Violence, Q22k*)
- e. Index of collective action*: index consisting of two components:
 - i. Indicator for talked with others about serious problem in community* (*CR Module, Social Inclusion, Q13a*)
 - ii. Indicator for took action with others about serious problem in community* (*CR Module, Social Inclusion, Q13b*)

3.5.6 Girls’ Economic Empowerment Outcomes

12. ***Index of economic empowerment:*** Index consists of three components:
- a. Indicator for had money she controls in past 12 months (*CR Module, Financial Inclusion and Economic Empowerment, Q1*)
 - b. Indicator for any savings for the future (*CR Module, Financial Inclusion and Economic Empowerment, Q2*)
 - c. Proportion of time spent in leisure and school on a typical weekday (including travel to and from, and studying) (*CR Module, Time Allocation, Q1-14*)

- d. Self-efficacy scale* (*CR Module, Mobility, Voice and Agency, Q12-21*)
13. **Index of economic aspirations:** Index consists of two components:
- a. Indicator for aspires to be employed in skilled or professional work when adult (*CR Module, Financial Inclusion and Economic Empowerment, Q3*)
 - b. Indicator for aspires to have employment or own a business when adult (*CR Module, Financial Inclusion and Economic Empowerment, Q3*)

3.5.7 Girls' Cross-Cutting Outcomes

14. **Index of gender equitable attitudes:** Index consists of two components:
- a. Global Early Adolescent Study's Index of Gender Stereotypical Traits¹¹: Index consists of eight components, including indicators for:
 - i. Disagrees with "Girls should avoid raising their voice to be lady like" (*CR Module, Education, Q25*)
 - ii. (Partially) agrees with "Boys should be able to show their feelings without fear of being teased" (*CR Module, Education, Q27*)
 - iii. Disagrees with "Girls are expected to be humble" (*CR Module, Education, Q26*)
 - iv. Disagrees with "It's important for boys to show they are tough even if they are nervous inside" (*CR Module, Education, Q33*)
 - v. Disagrees with "Boys who behave like girls are considered weak" (*CR Module, Education, Q32*)
 - vi. Disagrees with "Girls need their parents' protection more than boys" (*CR Module, Education, Q31*)
 - vii. Disagrees with "Boys should always defend themselves even if it means fighting" (*CR Module, Education, Q30*)
 - viii. Disagrees with "Boys should be raised tough so they can overcome any difficulty in life" (*CR Module, Education, Q29*)
 - b. Global Early Adolescent Study's Index of Gender Stereotypical Roles: Index consists of eight components, including indicators for:
 - i. (Partially) agrees with "Women should have the same chance to work outside of the home as men" (*CR Module, Financial Inclusion and Economic Empowerment, Q5*)

¹¹ For more information on the GEAS, see geastudy.org.

- ii. (Partially) agrees with “Girls and boys should share household tasks equally” (*CR Module, Time Allocation, Q17*)
- iii. Disagrees with “A woman’s most important role is to take care of her home and cook for her family” (*CR Module, Time Allocation, Q18*)
- iv. Disagrees with “A man should have the final word on decisions in his home” (*CR Module, Time Allocation, Q19*)
- v. Disagrees with “A woman should obey her husband in all things” (*CR Module, Sexual and Reproductive Health, Q27*)
- vi. Disagrees with “A boy should always have the final say about decisions with his girlfriend” (*CR Module, Sexual and Reproductive Health, Q20*)
- vii. Disagrees with “It is okay to tease a girl who acts like a boy” (*CR Module, Sexual and Reproductive Health, Q21*)
- viii. Disagrees with “It is okay to tease a boy who acts like a girl” (*CR Module, Sexual and Reproductive Health, Q22*)

15. ***Index of gender consciousness:*** Index consists of five components, including indicators for:

- a. (Partially) agrees with “Our culture makes it harder for girls to achieve their goals than boys” (*CR Module, Education, Q21*)
- b. (Partially) agrees with “I’m very aware of people’s reactions to my being a girl” (*CR Module, Sexual and Reproductive Health, Q23*)
- c. (Partially) agrees with “I think about how boys and girls roles differ from each other” (*CR Module, Sexual and Reproductive Health, Q24*)
- d. (Partially) agrees with “I think it is possible to change how people react to my being a girl” (*CR Module, Sexual and Reproductive Health, Q25*)

16. ***Index of supportive network:*** Index consists of three components, including: indicators for has female friend can trust and talk to, has male friend can trust and talk to, and has an adult can trust and talk to (*CR Module, Social Inclusion, Q4-5*)

17. ***Index of service knowledge*:*** Index consists of six components, including: indicators for knows a place where services could be received for alcohol and drug addiction (*CR Module, Health and Nutrition, Q25-25a*), mental health, pregnancy prevention, abortion/adoption, violence (*CR Module, Violence, Q23-23a*), injustice under the law

18. ***Index of service accessibility*:*** Index consists of seven components, including: indicators for believes an adolescent in his/her area could access services for alcohol and drug addiction (*CR*

Module, Health and Nutrition, Q25b), mental health, pregnancy prevention, abortion/adoption, violence (*CR Module, Violence, Q23b*), injustice under the law, financial¹²

19. **Knowledge index:** Index consists of fourteen components, including indicators for correct answers to the following:

- a. Do boys or girls usually reach puberty first? (Correct = “girls”) (*CR Module, Sexual and Reproductive Health, Q14*)
- b. Approximately how often does menstruation usually occur? By menstruation, I mean the bleeding girls get when they reach a certain age. (Correct = “once per month” or similar response) (*CR Module, Sexual and Reproductive Health, Q15*)
- c. True or False: The beginning of menstruation means that girls have the ability to become pregnant. (Correct = true) (*CR Module, Sexual and Reproductive Health, Q16*)
- d. True or False: Getting pregnant early in puberty can be bad for the health of the girl and the baby. (Correct = true) (*CR Module, Sexual and Reproductive Health, Q17*)
- e. Index for correct answers to the following: I am going to read you a list of foods. For each one, please tell me if that food has a lot of iron, which can help you grow and prevent anemia. Chicken? Eggs? Injera? Milk? (Correct = chicken, eggs, injera) (*CR Module, Health and Nutrition, Q17a-d*)
- f. How many main meals is it recommended that a girl/boy your age should eat in a day, to be healthy? (Correct = 3 or more) (*CR Module, Health and Nutrition, Q18*)
- g. At what age is it legal for women to get married in Ethiopia? (Correct = 18) (*CR Module, Marriage and Relationships, Q37-37a*)
- h. At what age is it legal for men to get married in Ethiopia? (Correct = 18) (*CR Module, Marriage and Relationships, Q38-38a*)
- i. Does the practice of FGM/C have risks? (Zero for those who have never heard of FGM/C; Correct = “yes”) (*CR Module, Sexual and Reproductive Health, Q43 and Q50*)
- j. Where can people seek help if they are experiencing violence? (Correct = coded response, 0 if DK a place) (*CR Module, Violence, Q23a*)
- k. If you wanted to save money, where could you keep that money, other than a safe space at home? (Correct = coded response) (*CR Module, Financial Inclusion and Economic Empowerment, Q4*)

¹² Questions related to financial service accessibility are with reference to the adolescent him/herself.

- l. Which of these things is *not* part of negotiation skills? (Correct = “don’t let the other person have a chance to say their point of view”) (*CR Module, Mobility, Voice and Agency, Q9*)
- m. True or False: Boys are biologically smarter than girls. (Correct = “false”) (*CR Module, Time Allocation, Q20*)
- n. True or False: Gender roles (by which I mean the tasks that girls and boys typically do, and the ways that girls and boys typically behave) cannot be changed. (Correct = “false”) (*CR Module, Time Allocation, Q21*)

3.5.8 Boys’ Outcomes

1. ***Index of gender equitable attitudes:*** (as described above)
2. ***Index of gender consciousness:*** (as described above)
3. ***Index of violence:*** composed of indicator for not perpetrating peer violence (as described above), and peer violence index (as described above)
4. ***Patient Health Questionnaire-9 Score*** (as described above)
5. ***Indicator for has a supportive adult*** (as described above)
6. ***Knowledge index*** (as described above)

3.5.9 Caregivers’ Outcomes¹³

1. ***Index of aspirations for adolescent:*** Index consists of four components (among caregivers for girls only); including:
 - a. Indicator for aspiring for girl to at least complete upper secondary school (*AF Module, Parenting, Q5*)
 - b. Expected age of marriage for adolescent (code “don’t know” values to median age, code “never” values to max age) (*AF Module, Parenting, Q30e*)
 - c. Indicator for aspiring for girl to work in skilled or professional employment as an adult (*AF Module, Parenting, Q32*)
 - d. Indicator for aspiring for girl to be employed or own a business as an adult (*AF Module, Parenting, Q32*)
2. ***Index of support for education:*** Index consists of four components (among caregivers for girls only; missing if adolescent not enrolled in school) including:
 - a. Indicator for assists with or supervises girl’s school work (*AF Module, Parenting, Q4a*)

¹³ Primary analysis will explore female caregiver responses only. We may include male caregiver responses in supplementary analysis.

- b. Indicator for reduces girl’s domestic chores so that she can attend school or do school work (*AF Module, Parenting, Q4b*)
 - c. Indicator for provides light so that girl can do school work at night (*AF Module, Parenting, Q4c*)
 - d. Indicator for provides a place for girl to study (*AF Module, Parenting, Q4d*)
 - e. Indicator for purchases uniforms for girl’s schooling (=1 if school does not require uniforms) (*AF Module, Parenting, Q4f*)
 - f. Indicator for pays for transport for girl to get to school (=1 if can walk/no transport required) (*AF Module, Parenting, Q4g*)
3. **Indicator for did not cut back food same for boys and girls** (all caregivers; past 4 weeks) (*AF Module, Health and Nutrition, Q28 and Q29*)
4. **Index of gender equitable attitudes:** Index consists of two components (among all caregivers):
- a. Subset of the Global Early Adolescent Study’s Index of Gender Stereotypical Traits¹⁴: Index consists of three components, including indicators for:
 - i. Disagrees with “Boys who behave like girls are considered weak” (*AF Module, Attitudes, Q26*)
 - ii. Disagrees with “Girls need their parents’ protection more than boys” (*AF Module, Attitudes, Q25*)
 - iii. Disagrees with “Boys should be raised tough so they can overcome any difficulty in life” (*AF Module, Attitudes, Q24*)
 - b. Global Early Adolescent Study’s Index of Gender Stereotypical Roles: Index consists of eight components, including indicators for:
 - i. (Partially) agrees with “Women should have the same chance to work outside of the home as men” (*AF Module, Attitudes, Q12*)
 - ii. (Partially) agrees with “Girls and boys should share household tasks equally” (*AF Module, Attitudes, Q4*)
 - iii. Disagrees with “A woman’s most important role is to take care of her home and cook for her family” (*AF Module, Attitudes, Q3*)
 - iv. Disagrees with “A man should have the final word on decisions in his home” (*AF Module, Attitudes, Q5*)
 - v. Disagrees with “A woman should obey her husband in all things” (*AF Module, Attitudes, Q6*)

¹⁴ For more information on the GEAS, see geastudy.org.

- vi. Disagrees with “A boy should always have the final say about decisions with his girlfriend” (*AF Module, Attitudes, Q31*)
 - vii. Disagrees with “It is okay to tease a girl who acts like a boy” (*AF Module, Attitudes, Q33*)
 - viii. Disagrees with “It is okay to tease a boy who acts like a girl” (*AF Module, Attitudes, Q32*)
5. ***Index of gender consciousness:*** Index consists of five components (among all caregivers), including indicators for:
- a. (Partially) agrees with “Our culture makes it harder for girls to achieve their goals than boys” (*AF Module, Attitudes, Q13*)
 - b. (Partially) agrees with “I’m very aware of people’s reactions to my being a woman (a man)” (*AF Module, Attitudes, Q21*)
 - c. (Partially) agrees with “I think about how men’s and women’s roles differ from each other” (*AF Module, Attitudes, Q23*)
 - d. (Partially) agrees with “I think it is possible to change how people react to my being a woman (a man)” (*AF Module, Attitudes, Q25*)

3.6 Data Collection and Processing

3.6.1 Survey Data Collection

The primary source of data used for this analysis will be survey data collected through face-to-face individual interviews with adolescents and their caregivers. In particular, we will employ three rounds of data collection in our analysis (a baseline and two post-programming follow-up rounds).

Baseline data was collected on the sample of adolescents and their caregivers during November 2017 through February 2018, by an Ethiopian data collection firm (Ethiopian Development Research Institute) using experienced survey enumerators who spoke the local languages of the study areas, and were well-trained in interviewing children. Baseline data collection survey instruments are available online (Baird *et al.*, 2019a, b, c, and d), and include:

- **Adult Female Module:** administered to the primary female caregiver of the core adolescent; collects information on the household (member roster, dwelling characteristics, asset ownership, agricultural and economic activities, decision-making, and productive capital), the adolescent core respondent (parent and sibling basic information, receipt of program benefits, and disability and functioning), and the adult female respondent herself (time use, access to credit and economic empowerment, use of information and communication technologies, health and nutrition, marriage

and fertility, attitudes, voice and agency, social interactions, experience of violence, parenting practices, and aspirations for the core adolescent)

- Core Respondent Module¹⁵: administered to the adolescent him/herself; collects information on education, time use, health and nutrition, psychosocial and mental health outcomes, mobility, voice and agency, social interactions, economic empowerment, use of information and communication technologies, marriage, sexual and reproductive health, experience of violence, and a Raven's Colored Progressive Matrices test

Baseline data collection also included survey interviews with a subset of approximately 20% of adult male caregivers, as well as key informants in the community (kebele leaders, school and health facility administrators) in order to collect basic information on the community and access to infrastructure and services there.

The baseline data has already been analyzed, primarily to understand patterns in outcomes across genders, regions, and adolescent disability status. This data has been used to inform the creation of the follow-up survey questionnaires.

Follow-up data collection is planned in two rounds, the first to take place during mid-November 2019 through March 2020, and the second to take place during mid-November 2020 through March 2021. Again, this data collection will entail face-to-face interviews with adolescents and their caregivers, using instruments similar to those discussed above. The survey enumeration team will attempt to contact all sample adolescents. Adolescents who have moved from their baseline location will be tracked, as long as they are still living within Ethiopia, using the two-stage tracking methodology described in Baird et al., (2016).

For secondary analysis, we will also receive adolescent group attendance information from Pathfinder and Care.

3.6.2 Data Quality Control and Processing

Quantitative data collection will be managed by an experienced data collection firm in Ethiopia with oversight from the GAGE Quantitative Research Team. Survey enumerators will be closely supervised to ensure the quality of their work. Daily telephonic communications between field supervisors and the in-country data collection partner will be maintained to monitor the progress and quality of the data collection.

Interviews will be conducted on tablet computers using the SurveyCTO software, with functionality that checks for internal validity and reasonableness of responses. Interviewers will ensure completion and accuracy of the survey before terminating each interview and transmitting the data back to the data

¹⁵ Note that the surveys were also administered to a larger sample of households that are not part of the current study (they are located in other areas of Ethiopia). Questions for older adolescents only are also included in these surveys.

collection firm. At the end of each day, the field supervisor will have de-briefing meetings with interviewers and discuss any problems or shortcomings, and they will take necessary actions in consultation with the in-country data collection firm and GAGE Quantitative Research Team members.

The in-country data collection firm and GAGE Quantitative Research Team members will review the data collected during field visits for completeness and consistency. Back checks will be conducted for a targeted subset of the sample, and will include re-interviews asking a subset of questions to respondents (for an interview lasting 5-10 minutes per respondent), generally conducted by supervisors. In addition, meetings will be arranged in the field to discuss progress and problems in the actual fieldwork and troubleshooting tips will be given if necessary.

Basic data cleaning and management will be performed by the GAGE Quantitative Research Team. This will primarily include verification of respondent identity and checks on uncoded answers to questions with coded response sets.

4 Empirical Analysis

4.1 Regression Specification

The main analysis will focus on the intention to treat program impacts in Amhara and Oromia across the set of primary and secondary outcomes listed in Section 3.5 at both 8 and 20 months post-program start (examining each round separately). ANCOVA regression analysis will be used to analyze the impact of the four treatment arms where we have baseline information on the outcome measure. We will utilize linear probability models in the case of binary outcomes. To answer our research questions, we will use two simple reduced form linear models:

$$y_{ic} = \alpha_1 + \beta_1 HS_c + \beta_2 AWH_c^{all} + X'_{ic} \beta_3 + \beta_4 y_{ic,0} + \varepsilon_{1,ic} \quad (1)$$

$$y_{ic} = \alpha_2 + \gamma_1 AWH_c + \gamma_2 Curr_c + \gamma_3 Asset_c + X'_{ic} \gamma_4 + \gamma_5 y_{ic,0} + \varepsilon_{2,ic} \quad (2)$$

where y_{ic} is the outcome of interest for individual i in community c , HS_c is a binary indicator for being in a *Her Spaces* community, AWH_c^{all} is a binary indicator for being in a community having any version of the *Act With Her* programming, AWH_c is a binary indicator for being in a community with the AWH-E main model, $Curr_c$ is a binary indicator for being in an AWH-E Simple community, and $Asset_c$ is an indicator for being in an AWH-E + Asset Transfers community. The first model would include all households, and the second model would exclude *Her Spaces* households. The standard errors ε_{ic} are clustered at the kebele level, which account for both the design effect of the kebele-level treatment and the heteroskedasticity inherent in the linear probability model. In the primary model, X_{ic} includes block indicators used in the randomization (where blocks combine woreda and marginalization status), adolescent age (in years) and an indicator for whether there were multiple eligible adolescents in the household. We will also provide

additional results that control for baseline household characteristics of interest, including: household size, indicator for household head literate, indicator for female headed household, household asset index, indicator for household receives PSNP benefits, and survey month indicators.¹⁶ Following McKenzie (2012), we will control for baseline measures of the outcome of interest (denoted by $y_{ic,0}$) when available. Gender and community specific sampling weights will be used to make the results representative of the target population in the study area.

We are primarily interested in five hypothesis tests. First, we seek to understand whether programming for girls only (*Her Spaces*), and whether the gender-synchronized, multilevel *Act With Her* programming for girls, boys, and parents, impact adolescent outcomes.

(i) $\beta_1 \neq 0$: comparing *Her Spaces* to the control group

(ii) $\beta_2 \neq 0$: comparing AWH (all variations) the control group

Then we seek to understand any added value of the latter programming relative to the former:

(iii) $\beta_1 \neq \beta_2$: comparing *Her Spaces* to AWH (all variations)

Finally, we are also interested in understanding the added value of specific components of the gender-transformative programming:

(iv) $\gamma_1 \neq \gamma_2$: comparing AWH-E to AWH-E Simple

(v) $\gamma_1 \neq \gamma_3$: comparing AWH-E to AWH-E + Asset Transfers

As a secondary analysis, we will conduct treatment on the treated analysis for the primary outcomes.

4.2 Missing Data

We describe our treatment of any differential attrition by treatment arm in Section 3.3.4 above. For individual outcomes with missing data (not from attrition), if missing is less than 20% we will simply ignore it. If 20% or more is missing across all arms, we will no longer utilize that variable. For missing data on covariates, we will follow Lin, Green and Coppock (2016). Observations with missing covariate values will be included in the regressions that estimate average treatment effects, as long as the outcome measure and treatment assignment are non-missing. We will use the following approach: (1) if no more than 10% of the covariate's values are missing, recode the missing values to the overall mean (do not use arm-specific means); (2) if more than 10% of the covariate's values are missing, include a missingness dummy as an additional covariate and recode the missing values to the overall mean.

¹⁶ There has been recent discussion over what the female household head measure captures (see for example the World Bank's Development Impact Blog of June 6, 2019). We will check the associations of this measure with other household characteristics before determining whether this measure will be included as a control. We will also explore the possibility of integrating machine learning for choice of control variables.

4.3 Handling of Outliers

We do not expect to have outliers among our predefined set of outcomes, all of which are composed of indicators and/or measures with well-defined ranges.

4.4 Baseline Balance Analysis

We will conduct baseline balance analysis across all primary and secondary outcomes that were measured at baseline as well as the following set of control variables: age, indicator for female adolescent, indicator for multiple eligible adolescents per household, household size, indicator for household head literate, indicator for household is female headed (if included as a control – see footnote above), household asset index, indicator for whether the household received PSNP benefits, and a set of indicators for survey month.

4.5 Multiple Outcome and Multiple Hypothesis Testing

To be conservative, given the number of hypothesis tests and primary outcomes we pre-specify here, we will construct FDR q-values, as described in Anderson (2008). This multiple-hypothesis testing will be conducted across the primary outcomes (19 for girls, 6 for boys, and 5 for caregivers) for each hypothesis and sample separately. These q-values will be presented alongside the standard per-comparison p-values.

4.6 Heterogeneous Effects

This evaluation is being conducted across two different regions in Ethiopia (Amhara and Oromia) and in areas with very different marginalization statuses. We expect it will be interesting to explore heterogeneity across these characteristics. Additional secondary analysis may explore other dimensions of interest.

5 References

Alexander-Scott, M., E. Bell and J. Holden (2016). “DFID Guidance Note: Shifting Social Norms to Tackle Violence Against Women and Girls (VAWG).” London: VAWG Helpdesk.

Baird, S., J. Hicks, N. Jones, J. Muz and the GAGE consortium (2019a). “Ethiopian baseline survey 2017/2018. Adult female module.” London: Gender and Adolescence: Global Evidence.

Baird, S., J. Hicks, N. Jones, J. Muz and the GAGE consortium (2019b). “Ethiopian baseline survey 2017/2018. Adult male module.” London: Gender and Adolescence: Global Evidence.

Baird, S., J. Hicks, N. Jones, J. Muz and the GAGE consortium (2019c). “Ethiopia baseline survey 2017/2018. Community module.” London: Gender and Adolescence: Global Evidence.

Baird, S., J. Hicks, N. Jones, J. Muz and the GAGE consortium (2019d). “Ethiopian baseline survey 2017/2018. Core respondent module.” London: Gender and Adolescence: Global Evidence.

- Baird, S., C. McIntosh and B. Özler (2011). “Cash or Condition: Evidence from a Cash Transfer Experiment”, *Quarterly Journal of Economics* 126: 1709-53.
- Baird, S., C. McIntosh and B. Özler (2019). “When the Money Runs Out: Do Cash Transfers Have Sustained Effects on Human Capital Accumulation?” *Journal of Development Economics*, 140: 169-185.
- Bandiera, O., N. Buehren, R. Burgess, M. Goldstein, S. Gulesci, I. Rasul and M. Sulaiman (2020). “Women’s Empowerment in Action: Evidence from a Randomized Control Trial in Africa.” *American Journal: Applied Economics*, 12(1): 210-59.
- Beaman, L., R. Chattopadhyay, E. Duflo, R. Pande and P. Topalova (2009). “Powerful Women: Does Exposure Reduce Bias?” *Quarterly Journal of Economics* 124 (4): 1497–1540.
- Blattman, C. and L. Ralston (2015). “Generating Employment in Poor and Fragile States: Evidence from Labor Market and Entrepreneurship Programs.”
- Buchmann, N., E. Field, R. Glennerster, S. Nazneen, S. Pimkina, and I. Sen (2018). “Power vs Money: Alternative Approaches to Reducing Child Marriage in Bangladesh, a Randomized Control Trial.” Unpublished manuscript, Abdul Latif Jameel Poverty Action Lab.
- Bundy, D., N. de Silva, S. Horton, G. Patton, L. Schultz et. al. (2017). “Child and Adolescent Health and Development: Realizing Neglected Potential.” In: *Disease Control Priorities* (third edition): Volume 8, *Child and Adolescent Health and Development*, edited by D. Bundy, N. de Silva, S. Horton, D. T. Jamison, G. Patton. Washington, DC: World Bank.
- Central Statistical Agency (CSA) [Ethiopia] and ICF (2016). “Ethiopia Demographic and Health Survey 2016: Key Indicators Report”. Addis Ababa, Ethiopia, and Rockville, Maryland, USA.
- Chakravarty, S., S. Haddock and I. Botea (2015). “Providing Out-of-School Girls with Skills: A Review of the Global Evidence.” World Bank Country Policy Brief. <https://openknowledge.worldbank.org/handle/10986/22799>.
- Chung, A. and R.N. Rimal (2016). “Social norms: a review.” *Review of Communication Research*, 4: 1–28.
- Corbin, C. and J. Heckman. (2016). “Capabilities and Skills.” *Journal of Human Development and Capabilities*, 17(3).
- Dahl, R., N.B. Allen, L. Wilbrecht, and A.B. Suleiman (2018). “Importance of investing in adolescence from a developmental science perspective.” *Nature*, 544: 441-450.
- Dhar, D., T. Jain, and S. Jayachandran (2020). “Reshaping Adolescents’ Gender Attitudes: Evidence from a School-Based Experiment in India.” *NBER Working Paper 25331*.
- Duflo, E (2012). “Women’s Empowerment and Economic Development.” *Journal of Economic Literature* 50: 1051-79.
- Edmeades, J, R. Hayes, and G. Gaynair (2014). “Improving the lives of married adolescent girls in Amhara, Ethiopia. A summary of the evidence.” Washington, DC: ICRW.

GAGE Consortium (2017). *Gender and adolescence: why understanding adolescent capabilities, change strategies and contexts matters*. London: Gender and Adolescence: Global Evidence.

Golla, A., A. Malhortra, R. Nanda, and R. Mehra (2011). “Understanding and Measuring Women’s Economic Empowerment: Definition, Framework and Indicators.” International Center for Research on Women (ICRW).

International Center for Research on Women (ICRW) (2016). “Child marriage facts and figures”. www.icrw.org/child-marriage-facts-and-figures.

Jayachandran, S. (2015) “The Roots of Gender Inequality in Developing Countries.” *Annual Review of Economics*, 7: 63-88.

Jensen, R., and E. Oster (2009). “The Power of TV: Cable Television and Women’s Status in India.” *The Quarterly Journal of Economics*, 124(3): 1057-1094.

Jones, N., B. Tefera, G. Emirie, W. Yadete, K. Gezahegne, K. Tilahun, K. Birhanu (2017). “Exploring Ethiopian adolescents’ gendered experiences and perspectives.” London: GAGE/ODI, Retrieved from <https://www.gage.odi.org/publications/exploring-ethiopian-adolescents-gendered-experiences-and-perspectives>.

Kabeer, N. (1999). “Resources, Agency, Achievements: Reflections on the Measurement of Women’s Empowerment.” *Development and Change*, 30(3): 435-464.

Kabeer, N. (2018). “Gender, livelihood capabilities and women’s economic empowerment: reviewing evidence over the life course.” London: Gender & Adolescence: Global Evidence, ODI.

Kato-Wallace, J., G. Barker, L. Sharafi, et al. (2016). “Adolescent boys and young men: Engaging them as supporters of gender equality and health and understanding their vulnerabilities.” Washington, DC, and New York: Promundo and UNFPA.

Kerner, B. (2016). “The Gender Roles, Equality and Transformations (GREAT) project in Northern Uganda”. Paper presented at Adolescence, Youth and Gender: Building Knowledge for Change, Oxford, 8–9 September.

La Ferrara, E., A. Chong, and S. Duryea (2012). “Soap operas and fertility: Evidence from Brazil.” *American Economic Journal: Applied Economics* 4(4): 1-31.

Marcus, R. and C. Harper (2015). *Social norms, gender norms and adolescent girls: A brief guide*. London: ODI.

McKenzie, D. (2012). “Beyond Baseline and Follow-up: The Case for More T in Experiments.” *Journal of Development Economics*, 99: 210-221.

McQueston, K., R. Silverman, and A. Glassman, A. (2013). “The Efficacy of Interventions to Reduce Adolescent Childbearing in Low- and Middle-Income Countries: A systematic review.” *Studies in Family Planning*, 44(4): 369–388.

Ministry of Education (2017). *Education Statistics 2008 E.C. (2015/16)*, Addis Ababa: Ministry of Education of the Federal Democratic Republic of Ethiopia.

Overseas Development Institute (2016). “Transforming the lives of Adolescent girls”. Retrieved from <https://www.odi.org/projects/2590-transforming-lives-adolescent-girls>.

Özler, B. (2016). “Keeping girls in school: a review of the global evidence.” Country Policy Brief. Washington, D.C. : World Bank Group. <http://documents.worldbank.org/curated/en/944241468188366610/Keeping-girls-in-school-a-review-of-the-global-evidence>.

Rahman, M., M. Kabir, M. Shahidullah (2007). “Participation of adolescents in household decision making processes in Bangladesh”, *Indian Journal of Community Medicine* 32(2): 123–27.

RISE (2018). “Research to Improve Student Learning Worldwide Will Study National Education Reform Package in Ethiopia.” Retrieved from <https://www.riseprogramme.org/news/research-improve-student-learning-worldwide-will-study-national-education-reform-package>.

Population Council (2020). “Adolescent Girls’ Empowerment”. Retrieved from <http://www.popcouncil.org/research/adolescent-girls-empowerment>.

Roba, A.C., K. Gabriel-Micheal, G.A. Zello et al. (2015). “A low pulse food intake may contribute to the poor nutritional status and low dietary intakes of adolescent girls in rural southern Ethiopia.” *Ecology of Food and Nutrition* 54(3): 240–54.

Institute for Reproductive Health, Georgetown University and Save the Children for the U.S. Agency for International Development (USAID) (2016). “Scaling Up Normative Change Interventions for Adolescent and Youth Sexual and Reproductive Health.” Washington, D.C.

Sheehan, P., K. Sweeny, B. Rasmussen, A. Wils, H. Friedman, J. Mahon, G. Patton, S. Sawyer, E. Howard, J. Symons, K. Stenberg, S. Chalasani, N. Maharaj, N. Reavley, H. Shi, M. Fridman, A. Welsh, E. Nsofor and L. Laski. (2017). “Building the foundations for sustainable development: a case for global investment in the capabilities of adolescents.” *The Lancet* 390(10104): 1792–1806.

Stavropoulou, M. and N. Gupta-Archer (2017a). “Adolescent girls’ capabilities in Ethiopia: The state of the evidence.” London: GAGE/ODI.

Stavropoulou, M. and N. Gupta-Archer (2017b). “Adolescent girls’ capabilities in Ethiopia: The state of the evidence on programme effectiveness.” London: GAGE/ODI.

Steinberg, L. (2015). *Age of opportunity: lessons from the new science of adolescence*. New York: Eamon Dolan/Mariner Books.

Thurnham, D. (2013). “Nutrition of adolescent girls in low- and middle-income countries.” *Sight and Life* 27(3): 26–37.

UNFPA (2014). United Nations Population Fund, Annual Report.

UNICEF (United Nations Children’s Fund) (2014). “A Statistical Snapshot of Violence Against Adolescent Girls”, UNICEF: New York.

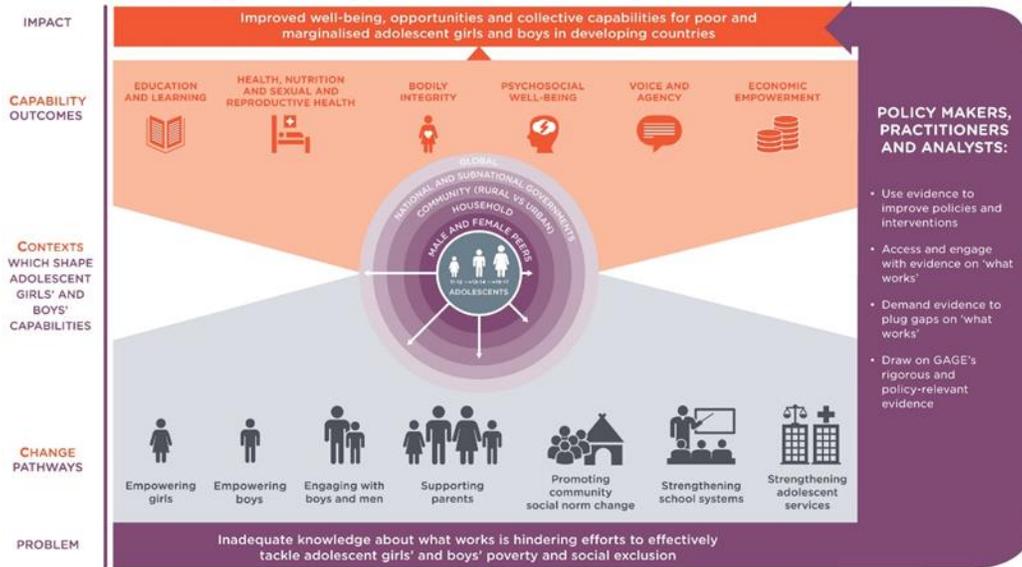
WHO (2014). “Health for the world’s adolescents: A second chance in the second decade of life”. Geneva: WHO.

World Bank (2016), “Indicators”. <http://data.worldbank.org/indicator>.

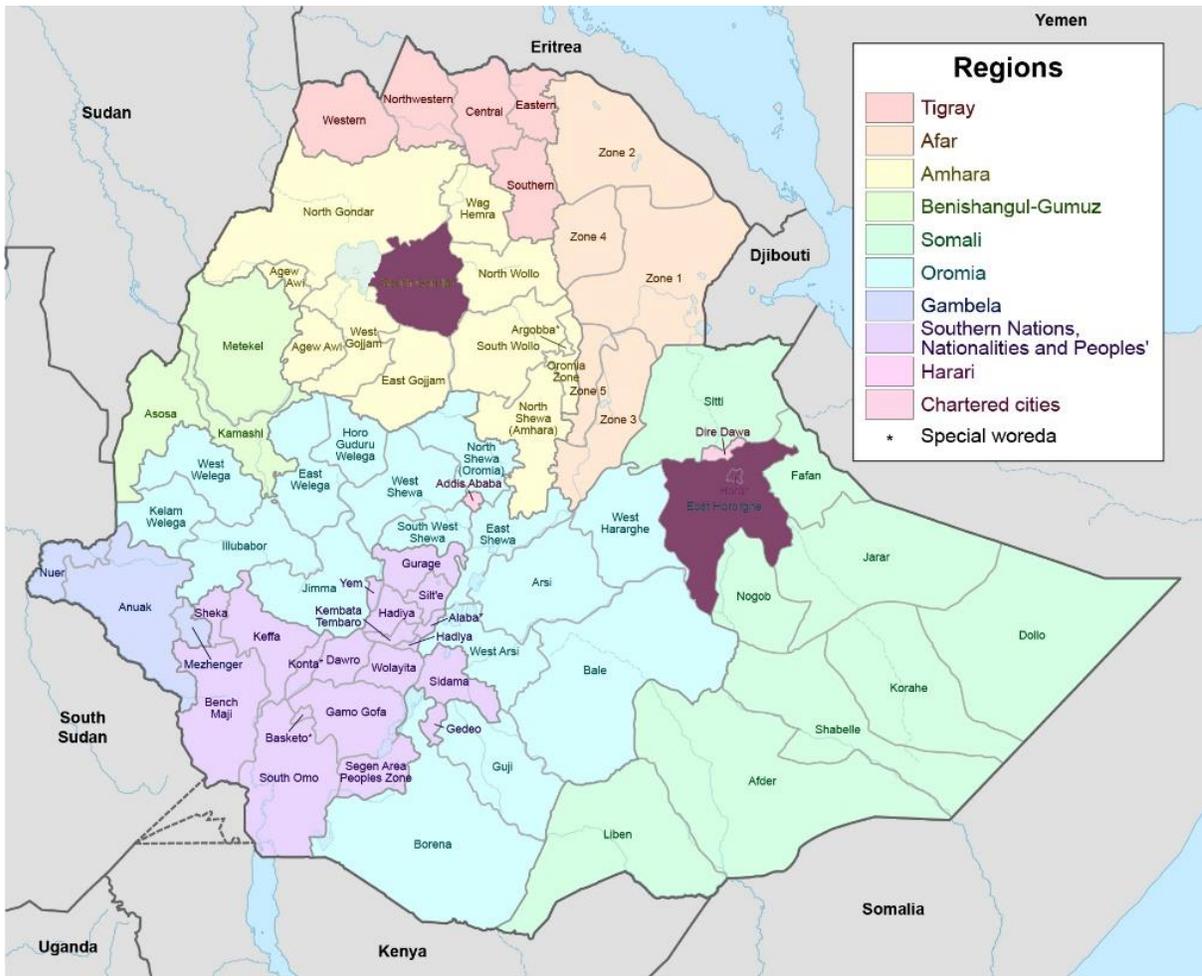
6 Appendices

6.1 Appendix A: Additional Figures

Appendix Figure A.1. GAGE conceptual framework



Appendix Figure A.2: Act with Her Locations (in dark purple)



Notes: Originally created from File:Ethiopia adm location map.svg by User:NordNordWest and modified to show GAGE research sites.

6.2 Appendix B: GAGE Ethiopia Listing Protocol: Rural Sites

We have selected several kebeles in Amhara and Oromia for the rural portion of the GAGE study. In order to select the households we will seek to interview for our main data collection activity in these kebeles, we will need to do a listing in each of them. **Your group of 4 enumerators and 1 supervisor will be assigned to particular kebeles** for this activity by the Field Coordinator.

You will not list the entire kebele, as it is too big. Instead, **your group will select a sub-kebele to work in according to the rule outlined below, and list all households in within one or more gots of this sub-kebele according to the process described below.**

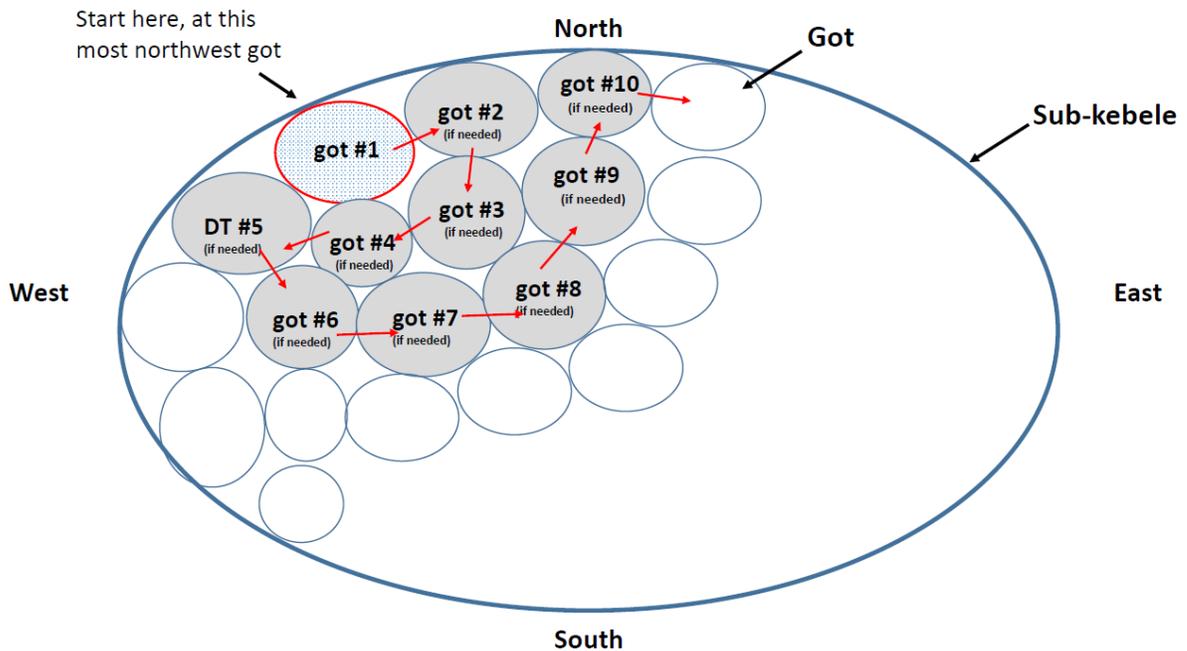
Here is the protocol for how we will do the listing within each kebele:

- When you arrive at the kebele that has been assigned to your group, discuss with the kebele leader (or other knowledgeable person) to make a list of all sub-kebeles in the kebele, along with the number of households living in each of these sub-kebeles. [Supervisor: You should include this list in the “Listing Information Document” and send to your field coordinator by the end of that day.]
- From this list, **choose the sub-kebele that has at least 200 households, and is closest to 200 households. This is the sub-kebele that you will be working in.**
 - For example, if sub-kebele A has 199 households, sub-kebele B has 205 households, and sub-kebele C has 210 households, you would select sub-kebele B. Note that sub-kebele A is not eligible for listing as it has fewer than 200 households.
 - If sub-kebele A has 199 households, and sub-kebele B has 500 households, you would select sub-kebele B. Again, sub-kebele A is not eligible for listing as it has fewer than 200 households. We describe a procedure below for listing within sub-kebele B such that you are not listing the entire sub-kebele.
- Within the sub-kebele you have selected, listing will proceed as follows:
 - **If you encounter a kebele in which all sub-kebele have fewer than 200 households, proceed by listing the sub-kebele with the most households. List that entire sub-kebele.** Inform the field coordinator of this issue immediately, and ensure that this message is quickly relayed to ODI/Quant HUB so that we can monitor. Our assumption is that this will not happen.
 - **If you have selected a sub-kebele that has more than two hundred households, listing will proceed as follows:**
 - Obtain local guides to assist your group.
 - Discuss with the guides to **locate the got that is in the far northwest corner of the selected sub-kebele.** This is the got that you will list first. List all households in this got, working with your guides to ensure you do not miss any households.
 - **If this got has at least 200 households, you will only list this first got, and then stop.**
 - **If this first got has fewer than 200 households, you will continue to the next got,** just to the north-east of the one you started in (as in the picture below).
 - Once you list the second got, again check to see whether your group has listed a total of at least 200 households. If yes, you may stop. If no, proceed to the next got, according to the picture below, snaking back to the south and west.
 - Once you list the third got, again check to see whether your group has listed a total of at least 200 households. If yes, you may stop. If no, proceed to the next got, according to the picture below.
 - [The movement in the picture is as follows: Continue in this manner, stopping at each got to list, until you have reached the western border again. Then, move east and north again, stopping at each got along the way. Continue to list each got you reach by snaking back and forth in this manner.]
 - **Continue to list one got at a time, in full and according to the movement in the picture below, until your group has listed a total of at least 200 households. If you reach 200 households in the middle of a got, you must complete the listing of all remaining households within that got before stopping.**

Points to remember:

- It is imperative that you always use the method outline above to select the sub-kebele and got(s) that you will list.
- You should **list all households within the selected got (or small sub-kebele)**, and then get additional information for the households that have eligible children.
- If you are working in a sub-kebele with fewer than 200 households, you will simply list the entire sub-kebele.
- If you reach 200 households in the middle of a got, you must complete the listing of all remaining households within that got before stopping.
- **Age verification process:**
 - **Recall that in the rural areas of Amhara and Oromia, we will only be interested in adolescents aged 10-12. The eligible birth years for these individuals is 1997-2000.**
 - When verifying age during the listing, we prefer that you use a birth certificate to verify age. If this is not available, request school report card to verify age. If also not available, check with the father or with an educated older sibling, as mothers are often not numerate.
- If we are unable to locate enough disabled children through the proposed listing strategy, we will need to speak to health extension workers, inclusive education teachers, and religious leaders in the sub-kebele in order to identify them. Any children found through this process should NOT be included in the listing data, but added in a supplementary document shared with the field coordinators. GAGE Quant Hub will let you know if we need to do this, depending on how the listing goes.
- ***Supervisors: please additionally collect the information contained in the “Listing Information Document” circulated separately, and pass along to the field coordinators.***

How to Choose Gots to List Within a Sub-kebele



6.3 Appendix C: Secondary Outcomes for Analysis

As in the main text, note that to construct indices, we employ the following procedure: a) for each index component, create a normalized measure by subtracting the control group mean and then dividing by the control group standard deviation; b) construct the index by calculating the raw mean across all normalized component variables, and c) create the normalized index value by subtracting the control group mean of the index and then dividing by the control group standard deviation of the index. We have signed outcomes such that higher values of indices indicate ‘better’ outcomes. * indicates items that will be measured at 20-months post-program start only.

C.1 Girls’ Education and Learning Secondary Outcomes

- ✓ Components of primary outcome index
- ✓ Indicator for no grade repetition in the last 12 months* (among those enrolled)
- ✓ Highest grade attended (zero for kindergarten or less) (*CR Module, Education, Q5b and Q6*)
- ✓ Indicator for ever enrolled in secondary school* (*CR Module, Education, Q5b and Q6*)
- ✓ Normalized score on educational tests (math/numeracy, Amharic reading, English reading)*

C.2 Girls’ Bodily Integrity Secondary Outcomes

- ✓ Components of primary outcome index
- ✓ Indicator for did not experience peer violence (as defined above) in last 12 months (*CR Module, Violence, Q10*)
- ✓ Indicator for did not witness or experience corporal punishment in school in past 12 months (among those enrolled) (*CR Module, Education, Q11, 11b, 12, and 12b*)
- ✓ Indicator for did not perpetrate peer violence in past 12 months (peer violence defined as described above) (*CR Module, Violence, Q11a-f*)
- ✓ Indicator for never married (*CR Module, Marriage and Relationships, Q1-2*)
- ✓ Age of first marriage (among married)* (*CR Module, Marriage and Relationships, Q5a*)

C.3 Girls’ Physical Health, Nutrition, and SRH Secondary Outcomes

- ✓ Components of primary outcome indices
- ✓ Body mass index (BMI)*
- ✓ Indicator for not ever pregnant (zero if never had sex)* (*CR Module, Sexual and Reproductive Health, Q30 and Q38*)
- ✓ Age at first pregnancy (among ever pregnant)* (*CR Module, Sexual and Reproductive Health, Q38a*)
- ✓ Desired fertility*

C.4 Girls' Psychosocial Wellbeing Secondary Outcomes

- ✓ Indicator for minimal depression (i.e., little sign of depression; PHQ-9 score ≤ 4)

C.5 Girls' Voice and Agency Secondary Outcomes

- ✓ Components of primary outcome indices (both subindices and their components)

C.6 Girls' Economic Empowerment Secondary Outcomes

- ✓ Components of primary outcome indices
- ✓ Indicator for in school, training, or studying (including traveling to and from; note: this will be a primary outcome in the 20-month follow-up analysis, replacing the 'proportion of time yesterday spent in leisure and school' measure) (*CR Module, Time Allocation, Q8-9*)
- ✓ Proportion of time spent in school, training, or studying (including traveling to and from; note: this will be a primary outcome in the 20-month follow-up analysis, replacing the 'proportion of time yesterday spent in leisure and school' measure) (*CR Module, Time Allocation, Q1-14*)
- ✓ Proportion of time in paid work*
- ✓ Indicator for any paid work in past 12 months* (*CR Module, Paid Work, Q1*)
- ✓ Wages in the past seven days* (*CR Module, Paid Work, Q1, 2, 4*)

C.7 Girls' Cross-Cutting Secondary Outcomes¹⁷

- ✓ Components of all primary outcomes indices
- ✓ Index of gendered attitudes toward education: Index of five components, including:
 - o. Disagrees with "If a family can afford for one child to go to secondary school, it should be the boy only" (*CR Module, Education, Q22*)
 - p. Disagrees with "Only boys should learn about science, technology, and math" (*CR Module, Education, Q23*)
 - q. Disagrees with "Girls should be sent to school only if they are not needed to help at home" (*CR Module, Education, Q24*)
 - r. (Partially) agrees with "A girl's marriage can wait until she has completed senior secondary school" (*CR Module, Sexual and Reproductive Health, Q28*)
 - s. Disagrees with "It is appropriate for parents to take boys out of school for work" (*CR Module, Education, Q28*)
- ✓ Index of attitudes toward violence: Index of three components, including :

¹⁷ We note that the GAGE survey asked a number of additional attitude questions on school violence, politics, and female empowerment. We intend to explore these questions in supplementary analysis.

- a. Disagrees with “It is acceptable for a man to hit or beat his wife in order to mold or control her behavior” (*CR Module, Violence, Q1*)
- b. Disagrees with “A man using violence against his wife is a private matter that should not be discussed outside the couple” (*CR Module, Violence, Q2*)
- c. Disagrees with “A woman should tolerate violence to keep her family together.” (*CR Module, Violence, Q3*)

C.8 Boys’ Secondary Outcomes

- ✓ Components of all primary outcome indices

C.9 Caregivers’ Secondary Outcomes

- ✓ Components of all primary outcome indices

7 Administrative Information

7.1 Funding

The evaluation proposed here is supported by the UK Department for International Development on behalf of UK Aid. The programming implementation described here is supported by the Bill & Melinda Gates Foundation, Seattle, Washington [grant number OPP1181670].

7.2 Institutional Review Board

The study design was approved by the George Washington University Committee on Human Research, Institutional Review Board (071721), the ODI Research Ethics Committee (02438), and the Addis Ababa University College of Health Sciences Institutional Review Board (113/17/Ext). A Memorandum of Understanding (MoU) was signed with the Ethiopia Federal Ministry of Women, Children and Youth. The study design was also approved by different regional-level bodies, including the Ethical Committee of the Amhara Region Bureau of Health (Amhara Region), the Ethical Committee of the Oromia Region Bureau of Health (Oromia Region), and Semera University Ethical Board (Afar Region).

7.3 Declaration of interest

Courtney L. Boudreau and Tricia Petruney are employed by Pathfinder International as the Monitoring, Evaluation, and Learning Manager and the Project Director, respectively, for the *Act With Her* project. The other authors have no conflicts of interest to declare.

7.4 Acknowledgements

We gratefully acknowledge comments on this protocol from Laura Hahn at the Bill & Melinda Gates Foundation, and anonymous external reviewers.