



Review article

A Systematic Review of Adolescent Girl Program Implementation in Low- and Middle-Income Countries: Evidence Gaps and Insights

Nicole A. Haberland, M.P.H.^{a,*}, Katharine J. McCarthy, M.P.H.^a, and Martha Brady, M.S.^b^a Population Council, New York, New York^b PATH, Washington, District of Columbia

Article history: Received June 30, 2017; Accepted November 9, 2017

Keywords: Adolescent girls; Programs; Intervention; Low- and middle-income countries; Implementation science; Operations research; Effectiveness

A B S T R A C T

Increasing attention to adolescent girls has generated an abundance of programs and a growing body of research on adolescent girls in low- and middle-income countries. Despite this, questions remain about what implementation approaches in program design are most effective, hindering efficient resource allocation, program scale-up, and replication across settings. To address these questions, we conducted a systematic review to identify lessons learned and gaps in the evidence base. We searched four electronic databases to identify studies published between 1990 and 2014 that evaluated health, social, and/or economic development programs targeting adolescent girls in low- and middle-income countries. Seventy-seven (77) studies meeting specified criteria were identified, of which 19 presented results that allowed conclusions relevant to implementation science. Studies examining the following questions were assessed: To what extent, if any, do multicomponent interventions (as opposed to single-component interventions) improve outcomes for girls? What is the added value of involving actors in addition to the girl herself such as parents, guardians, husbands (i.e., multilevel interventions)? What is the threshold proportion of girls who need to participate in a program to bring about normative and behavior changes at the community level? Is a greater level of program exposure associated with greater programmatic benefit for girls? Can supplemental “booster” activities extend the benefits of a program after it ends? We found evidence to support associations between multicomponent (vs. single component) programs, and longer program exposure (vs. less program exposure), with more favorable outcomes for girls, although both conclusions include methodological limitations. Overall, few studies assessed boosters or program saturation, and evidence on multilevel versus single-level programs was inconclusive. Few studies assessed implementation science questions by design, exposing large gaps in the evidence base. We call for future research to explicitly test such implementation science questions to inform more effective use of resources and to improve outcomes for girls.

© 2017 Society for Adolescent Health and Medicine. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

IMPLICATIONS AND
CONTRIBUTIONS

Evidence from girl-centered programs in low- and middle-income countries suggests longer program exposure and multicomponent (vs. single component) programs may be more effective. Substantial evidence gaps in program implementation are identified. How to improve program design to maximize outcomes for girls is the next-generation question.

Conflicts of Interest: The authors declare no competing interests.

* Address correspondence to: Nicole Haberland, M.P.H., One Dag Hammarskjöld Plaza, New York, NY 10017.

E-mail address: nhaberland@popcouncil.org (N.A. Haberland).

The vast majority (86%) of the world's adolescents ages 10–24 live in low- and middle-income countries (LMICs) [1]. Within resource-limited contexts, girls face distinct challenges across multiple health, social, and economic domains. Among adolescents

ages 15–19, approximately two thirds of new HIV infections are acquired by girls [2]. Further, globally, 11% of births occur among adolescents ages 15–19, and nearly all (95%) of these occur in LMIC settings [3]. Complications during pregnancy and childbirth are a leading cause of death among girls ages 15–24 [4]. The factors that underlie these adverse health outcomes are largely social and economic, reflecting societies' general devaluation of girls and harmful gender norms. For example, child marriage, or marriage before the age of 18, affects an estimated 15 million girls globally every year [5]; more than 85% of girls in low-income countries never complete secondary school [6], and intimate partner violence affects an estimated 29.4% of ever-partnered girls ages 15–19 worldwide [7]. Not only do these practices undermine girls' rights, agency, and current sexual and reproductive health, but such practices also limit their economic opportunities and affect their health and well-being into adulthood [8–12].

These pervasive needs have led to increasing interest in programs that target girls in LMIC [13–15]. Recognition of the interrelated nature of girls' social, economic, and sexual and reproductive health vulnerabilities has increasingly led to program innovations that aim to directly address the gender inequalities adolescent girls face. These include multicomponent programs that combine different interventions—such as life skills education and savings accounts—that aim to redress inequalities faced by girls by building girls' protective assets and thus improving the likelihood of positive health and development outcomes. For example, a theory of change may posit that a girl who is able to increase her economic assets is able to increase her relative power and will be better able to act on information about HIV prevention and better positioned to negotiate condom use than a girl who only receives information about HIV. Similarly, multilevel programs that reach not just the girl but those who act as her gatekeepers—such as parents or husbands—may posit that it is more likely for an intervention directed at girls to succeed if the enabling environment is supportive of change. Such programs might include, for example, economic incentives for girls to stay in school in addition to activities with parents/guardians to increase their support for girls' education.

However, questions remain about whether such combined programs do, in fact, perform better than programs with a single component or a single level. Questions about other implementation design elements, such as optimal program length, also persist. See [Figure 1](#) for definitions of key terms used in this paper.

Given the magnitude of investments in girls programming globally, as well as the implications for achieving current development agendas, such as the 2030 Sustainable Development Goals, greater understanding of what intervention designs are most effective in promoting positive outcomes for adolescent girls in LMIC is urgently warranted. Several recent reviews of adolescent programs have examined the effectiveness of programs for girls (e.g., Hardee et al., 2014 [HIV]; Hennegan and Montgomery, 2016 [menstrual pads for education]) or programs directed at outcomes that disproportionately impact female adolescents (e.g., Kalamar et al. 2016 [child marriage]; Hindin et al., 2016 [unintended repeat pregnancy]; Lundgren and Amin 2015 [intimate partner violence]) [16–20]. These reviews examine the evidence for effectiveness of different programs—such as school-based life skills, conditional cash transfers (CCTs), or youth friendly services. Yet, to our knowledge, no reviews have explicitly explored implementation questions—such as how long a girl needs to be in a program, or what proportion of girls in a community participating can generate a tipping point for sustained change—and the

relative impact of such program design choices. As girl-centered programs are considered for replication, expansion, and scale, or new program ideas are innovated for pilots, it is critical that we know what works for girls, and equally important, what does not, to guide investment of finite resources.

We conducted a systematic review of the published and gray literature to identify evidence gaps and what implementation science lessons can be learned about fielding successful adolescent girl-centered programs in LMIC settings. Specifically, we sought to examine studies that explicitly tested variations in intervention design or structure to assess which aspects are most likely to lead to improvements in girls' health, social, or economic outcomes. The following questions guided the review:

- To what extent, if any, do multicomponent interventions (as opposed to single component interventions) improve outcomes for girls?
- What is the added value of involving actors in addition to the girl herself such as parents, guardians, husbands (i.e., multi-level interventions)?
- What is the threshold proportion of girls who need to participate in a program to bring about normative and behavior changes at the community level?
- Is a greater level of program exposure associated with greater programmatic benefit for girls?
- Can supplemental “booster” activities extend the benefits of a program after it ends?

Methods

Data sources

Studies were identified using a keyword search of four electronic databases: PubMed, the Cumulative Index to Nursing and Allied Health Literature (CINAHL), Economic Literature, and Sociological Abstracts. Search strings included key terms related to the study population, program elements, and outcomes in LMIC. We focused on girl's programs in three broad sectors: health (i.e., sexual and reproductive health, HIV and sexually transmitted infections [STIs]), social (i.e., education, violence, empowerment), and economic. The full list of the search terms is listed in [Supplemental File 1](#). The initial search yielded a total of 44,460 studies, including peer-reviewed articles, book chapters, working papers, and program briefs (henceforth “studies”). An additional 30 studies were identified from a web-based search of non-governmental organizations (NGOs) that implement programs to improve the health and well-being of adolescent girls. The PRISMA protocol guided the review [21].

Inclusion/exclusion criteria

We considered studies that were (1) published between January 1, 1990, and April 30, 2014, in English; (2) targeted adolescent girls and young women aged 10–24 residing in LMIC; (3) examined changes in knowledge, attitudes, behavior, and/or status (such as pregnancy, employment, grade attainment, marriage) after exposure to a health, social, and/or economic intervention; and (4) reported quantitative outcomes either adjusted for or disaggregated by gender and age group of interest.

Studies were excluded if (1) adolescent girls did not comprise at least 50% of the sample, (2) there were fewer than 100

Definition of Terms

Implementation science – “the study of methods to promote the adoption and integration of evidence-based practices, interventions and policies into routine health care and public health settings.”¹ Implementation science seeks to “to improve program effectiveness and optimize efficiency” and facilitate “making evidence-based choices between competing or combined interventions and improving the delivery of effective and cost effective programs.”² It can also be used to “support the process of re-iterative refinement needed for successful adaptation.”³

Multicomponent – programs that include more than one type of intervention for participants; for example, a program that includes both life skills education and livelihoods training, or a program where participants both receive academic tutoring and participate in empowerment groups. Such programs have also been called combined (versus single-focus) programs.

Multi-level – programs that reach not just the main beneficiary, but also include activities for actors who directly or indirectly affect their lives and well-being. Using the example of adolescent girls, this could include parents/guardians, husbands, siblings, teachers, etc.

Booster – supplemental activities implemented with participants after the end of the main intervention with the aim of sustaining program effects.

Saturation – the proportion of individuals in a community eligible to participate (based on program specific-criteria) who actually participated in the intervention.

Exposure – amount of a program’s intended content that was received by participants.

Durability – the sustainability of program effects over time.

¹ National Institutes of Health, Implementation science information and resources. Washington DC: Fogarty International Center, 2017.

² Padian N, Holmes C, McCoy S, Lyerla R, Bouey P, and Goosby E. Implementation science for the US President’s Emergency Plan for AIDS Relief (PEPFAR). *J Acquir Immune Defic Syndr*, 2011;56(3): 199-203.

³ Peters H, Tran NT, Adam T. Implementation research in health: a practical guide. Alliance for Health Policy and Systems Research, Geneva: World Health Organization, 2013.

Figure 1. Definition of terms.

girls enrolled, (3) there was no comparison group in the evaluation or intervention exposure was based on recall, and (4) the program was not “girl-centered.” Programs were considered girl-centered if they met at least two of four conditions: (1) the

program explicitly targeted or intended to reach adolescent girls, (2) included content sensitive to girls’ unique situation and/or needs, (3) intended to address girl-specific needs or vulnerabilities, and (4) evaluated and discussed outcomes for girls. In cases

where it was unclear whether a program is qualified, the authors (M.B., N.H., K.J.M.) met to reach group consensus.

Data extraction

Three independent reviewers extracted basic study information from 77 eligible studies, including citation, program name, study design, population, setting, and sample size. Data were also extracted on the program objective, intervention content, and program attributes including whether the program was multi- or single component (i.e., did girls receive a single intervention such as life skills, or multiple interventions such as life skills and job training); multi- or single level (i.e., did the intervention only target girls or did the program also reach others, such as husbands or parents, that influence girls' lives); measured community saturation level (i.e., what proportion of eligible girls in the community were reached); measured participant exposure to the intervention (i.e., the intended program dosage and the amount received in practice); or whether it included a booster component. For each attribute, reviewers indicated whether the relative effect of the enhanced (i.e., multicomponent/level or higher saturation/exposure) versus "basic" version (single component/level or low saturation/exposure) was assessed by the study. Outcomes of interest included those related to knowledge, attitudes/beliefs/norms, self-efficacy/agency, behaviors, or health or status outcomes (e.g., STI, school enrollment, child marriage) at each follow-up assessment (if multiple). The extraction database is available upon request by the first author.

Data synthesis

We analyzed the subgroup of studies that compared programmatic elements of interest (e.g., multicomponent program performance vs. single component program performance) by design *within* the study ($N = 19$). Given variation across studies in the types of programs being evaluated, delivery structure, and participants, between-study comparisons were not considered informative, nor was it feasible to conduct a meta-analysis. Analyzed studies either: (1) directly compared study arms using statistical tests for difference ($p < .05$); or, if no statistical difference between arms was reported, we considered descriptive assessments such as (2) a greater number of significant intended effects in the program arm of interest relative to the baseline or control condition than in the comparison arm, or (3) a larger magnitude of effect produced by the program arm of interest relative to the comparison arm. We note the level of evidence provided by each study (Tables 1–3).

Methodological quality assessment

Study quality was assessed in two steps. First, we assigned an initial rating based on the type of study design where randomized controlled trials were considered high-quality, quasi-experimental designs, or a pre- and postintervention assessment with a comparison group was considered medium quality, and cross-sectional, control-comparison data as low quality [22]. Second, we adjusted the initial rating based on whether there were other sources of unaddressed bias in the study. We considered sources of bias listed in the Effective Public Health Practice quality assessment tool, including selection bias, potential for unadjusted confounding, differential attrition, potential for contamination between study groups, and appropriate use of sta-

tistical methods [23]. If any of the listed sources of bias were unaddressed in a randomized controlled trial study, for example, the "high" rating was downgraded to "moderate," or a "moderate" pre- and postcomparison study was downgraded to "low."

Results

Descriptive findings

Description of identified programs. We identified 77 studies evaluating a "girl-centered" health, social, or economic development program in LMIC settings, representing 61 distinct programs (Figure 2). The majority (70%) of studies took place in either sub-Saharan Africa or South Asia, 15% were in Latin America and the Caribbean, and less than 10% took place in the Middle East and North Africa, Central Asia, or East Asia and the Pacific, respectively. About one quarter of studies (28%) were high quality, most were moderate quality (52%) and 21% were low quality. In terms of program objectives (multiple objectives were possible), more than half (52%) of programs included a focus on sexual and reproductive health, followed by HIV (33%), girls' empowerment or leadership (30%), education (30%), and development of economic assets (25%). Comparatively fewer programs emphasized enhancing social support (16%), maternal and child health (15%), child marriage (11%), nutrition (11%), or sought to reduce violence (8%).

Overall findings

Forty-nine studies and 41 programs examined changes in behaviors; 76% of these programs identified a significant improvement in behavioral indicators among adolescent girls.

Fifty-three studies and 48 programs examined impact on health status or other outcomes (i.e., educational attainment or enrollment, child marriage); of these, 92% of programs identified a statistically significant beneficial effect among adolescent girls.

Implementation science findings

In total, 19 studies, evaluating 18 programs, provided sufficient information to address implementation science questions. Two studies provided information on more than one type of program attribute. Of the included studies, 6 were high quality, 11 were moderate, and 2 were low quality. We consider program performance by study quality in the sections that follow.

Multicomponent. Multicomponent programs were defined as programs that included more than one type of intervention for participating girls. Of the 61 programs, about half (30) were multicomponent and the remainder (31) were single component programs. Eight studies compared a single component program (e.g., life skills) with a multicomponent version (e.g., life skills plus savings accounts) in their study design (Table 1). In total, five studies [24,25,27,28,30] found stronger effects in the multicomponent arm, and three [26,29,31] found that the multicomponent arm did not perform better than the single component arm. Of the five studies that found stronger effects in the multicomponent arm, four (all medium- and high-quality studies) were based on changes in behavior or impact related to work, violence, school, or marriage; none were based solely on changes in knowledge, one (a low-quality study) was based solely on

Table 1Adolescent girl programs in low- and middle-income countries that assessed multicomponent^a versus single component intervention arms (N = 8)

Program	Citation	Intervention arms ^b	Study quality	Significant outcomes for girls (Intervention arms vs. Baseline/control) ^c	Does multicomponent arm perform better? ^d (Direct comparison of intervention arms or across arm evidence)
Kishori Abhijan, Bangladesh	Amin and Suran, 2005 [24]	1. APON (Education on health + legal rights) (S) 2. APON + microcredit to qualified members only (M1) 3. CMES (Education on health and legal rights + microcredit) (M2)	Medium quality	Follow-up at 24 mo Delayed marriage (ns) Dowry payment (ns) School retention (ns) Paid work (S+, M2+)	Yes^g Delayed marriage θ^f Dowry payment θ^f School retention θ^f Paid work \uparrow^g <ul style="list-style-type: none"> There were no statistical differences in delayed marriage, dowry payment, or school retention in S, M1, or M2 compared with matched nonparticipants Paid work increased in both S and M2, but M2 participants had a larger magnitude of effect
Safe and Smart Savings, Kenya & Uganda	Austrian and Muthengi, 2013 [25]	1. Savings accounts (S) 2. Safe spaces girls group + savings accounts (M)	Medium quality	Follow-up at 12 mo Attitudes that condone gender-based violence (GBV) (M–) Experiences of GBV (S+) Sexual/reproductive health (SRH) knowledge (M+) Ever HIV tested (ns)	Yes^f Attitudes that condone GBV \downarrow^f Experiences of GBV θ^f SRH knowledge \uparrow^f Ever HIV tested θ^f <ul style="list-style-type: none"> Girls in M arm had a greater number of improved outcomes for attitudes toward GBV, SRH knowledge, and no increase in experiences of GBV (vs. baseline) compared with girls in the S arm (vs. baseline)
Training and wage subsidy intervention, Jordan	Groh et al., 2012 [26]	1. Job voucher (S1) 2. Employability training (S2) 3. Job voucher + employability training (M)	High quality	Follow-up at 6 and 14 mo Current employment (S1+) Ever employed (S1+, S2+) Months employed (S1+, S2+) Hours worked last week (S1) Work income (S1) Life evaluation (current)* (S1+, M2–) Life evaluation (future)* (S2+) Mental health* (S2+) Mobility* (S1–, M2+) Empowerment* (ns) Marriage* (ns) *Examined at time 2 only	No^f <ul style="list-style-type: none"> Across arm evidence (S1 vs. control, S2 vs. control, M vs. control) does not show that M leads to a greater number of intended outcomes than S1 or S2 (see previous column) at either follow-up time
Siyakha Nentsha, South Africa	Hallman and Roca, 2011 [27]	1. Basic (social + health) education (S) 2. Basic education + financial education (M)	High quality	Follow-up at 18 mo Know where to get condoms (S+, M+) Know social grant requirements (S+, M+) Improved budget and planning skills (S+, M+) Attempts to open bank account (S+, M+) Saving behavior (ns) Remain sexually abstinent (ns) Fewer number of sexual partners (ns) Undertake income generating activity (ns) Self-esteem (S+, M+) Confidence in ability to get a condom (S+, M+) Perceived social inclusion (M+) Obtain birth certificate (M+)	Yes^e Perceived social inclusion \uparrow^e Obtained birth certificate \uparrow^e <ul style="list-style-type: none"> Girls in the M arm felt greater social inclusion and were more likely to have obtained a national birth certificate than girls in the S arm Other improvements were observed in both the S and the M arms relative to the control arm (see previous column)

(continued on next page)

Table 1
Continued

Program	Citation	Intervention arms ^b	Study quality	Significant outcomes for girls (Intervention arms vs. Baseline/control) ^c	Does multicomponent arm perform better? ^d (Direct comparison of intervention arms or across arm evidence)
Supporting adolescent orphan girls to stay in school, Zimbabwe	Hallfors et al., 2011 [28]	1. School feeding program (S) 2. School feeding program + school support (school fees and supplies paid; school helper to meet attendance requirement) (M)	High quality	M vs. S over time (baseline, 12 mo, 24 mo) School attendance (M+) Perception adults are caring (ns) Educational aspirations (ns) Future expectations about school completion (M+) Gender equitable attitudes (M#) Wife beating endorsement (ns) Think it is OK to ask husband to use condom (ns) Think it is not okay to have sex as an adolescent (ns) Waiting for sex until marriage/ because of values (ns) Waiting for sex because of consequences (M+) Ever sex (ns) School dropout* (M–) Delayed marriage* (M+) <i>*By study condition at end of study only</i> <i>#Marginally significant at p = .07</i>	Yes^e School attendance ↑ ^e Perception adults are caring θ ^e Educational aspirations θ ^e Future expectations about school completion ↑ ^e Gender equitable attitudes ↑ ^{#e} Wife beating endorsement θ ^e Think it is OK to ask husband to use condom θ ^e Think it is not okay to have sex as an adolescent θ ^e Waiting for sex until marriage/because of values θ ^e Waiting for sex because of consequences ↑ ^e Ever sex θ ^e School dropout* ↓ ^e Delayed marriage* ↑ ^e • Direct comparison of M vs. S shows the full intervention results in several intended effects sustained over time, not observed in the single component arm <i>*Marginally significant at p = .07</i>
Sanitary pad and puberty education program, Ghana	Montgomery et al., 2012 [29]	1. Puberty education (S) 2. Education + menstrual pads (M)	Medium quality	Follow-up at 3 and 5 mo School attendance (S1+, S2+, M1+, M2+)	No^g School attendance θ ^g • At both follow-up times, improvements in school attendance in S and M relative to baseline were similar in magnitude
Entre Amigas, Nicaragua	Peña et al., 2008 [30]	1. Peer groups (S) 2. Peer groups + mothers (M1) 3. Peer groups + mothers + soap opera (M2)	Low quality	Follow-up at 15 mo Self-esteem (S+, M1+, M2+) Gender equitable attitudes (S+, M1+, M2+)	Yes^e Self-esteem θ ^e Gender equitable attitudes ↑ ^e • Girls exposed to all 3 components (M2) had a significantly larger increase in gender equitable attitudes than those in M1 or S alone, providing evidence to support enhanced effect of multicomponent arm • Self-esteem increased with participation in S, regardless of exposure to M1 or M2, suggesting no enhanced multicomponent benefit for this outcome
Kishoree Kontha (Adolescent Girls' Voices), Bangladesh	Scales et al., 2013 [31]	1. Basic (SRH/life-skills, literacy) support (S1) 2. Livelihoods (basic + financial education) (S2) 3. Both (M)	High quality	Follow-up between 4 and 9 mo Developmental assets (S1+, S2+, M+)	No^g Developmental assets θ ^g All groups made significant improvement relative to baseline. Girls in the M arm did not show the largest magnitude of improvement, which was observed in the S1 arm

^a Multicomponent intervention refers to more than one type of intervention for the same participants.^b Program components indicated as follows: S = single component arm (S1 and S2 used to note multiple single component arms, if applicable); M = multicomponent arm (M1 and M2 refer to more than one multicomponent arm, if applicable).^c Outcomes: Subscripts represent multiple follow-up times (if applicable). S+/M+ = significant positive effect for girls in single component arm versus control/baseline condition or multicomponent arm versus control/baseline condition, unless noted otherwise; S–/M– = significant negative effect for girls in single component arm versus control/baseline condition or multicomponent arm versus control/baseline condition; ns = indicates no difference for girls in any comparison group at alpha=.05 level.^d Yes = multicomponent arm outperforms single component arm; No = multicomponent arm does not outperform single component arm. ↑ = positive effect; θ = no difference; ↓ = negative effect. Multicomponent arm was considered to outperform a single component arm within studies using the following levels of evidence for at least one outcome or at least one postintervention time point (if multiple).^e Direct comparison of multicomponent versus single component using significance test ($p < .05$).^f Greater number of intended significant outcomes in multicomponent versus control (or baseline) than in single component versus control (or baseline).^g Larger magnitude of effect in multicomponent versus control (or baseline) relative to single component arm versus control (or baseline).

Table 2Adolescent girl programs in low- and middle-income countries that assessed multilevel^a versus single-level component intervention arms (N = 5)

Program	Citation	Intervention arms ^b	Study quality	Significant outcomes for girls (Intervention arms vs. Baseline/control) ^c	Does multilevel arm perform better? ^d (Direct comparison of intervention arms or across arm evidence)
Exploring the World of Adolescents (EWA), EWA with parents (EWA+), Vietnam	Pham et al., 2012 [32]	1. SRH education + gender content (EWA curriculum) (S) 2. SRH + gender content + parent education (EWA+) (M)	High quality	Follow-up at 3, 6, and 12 mo Knowledge Pregnancy/contraceptive (S ₁₊ , S ₂₊ , M ₁₊ , M ₂₊ , M ₃₊) STIs (S ₁₊ , S ₂₊ , M ₁₊ , M ₂₊) HIV (S ₁₊ , S ₂₊ , M ₁₊ , M ₂₊ , M ₃₊) Attitudes toward risk and protective behaviors Extrinsic rewards (S ₁₊ , S ₂₊ , S ₃₊ , M ₁₊ , M ₂₊ , M ₃₊) Intrinsic rewards (S ₁₊ , S ₂₊ , S ₃₊ , M ₂₊) Perceived severity: pregnancy (M ₂₊) Perceived severity: HIV/AIDS (M ₂₊) Perceived vulnerability: sex (S ₁₊ , S ₃₊) Perceived vulnerability: HIV/AIDS (S ₁₊ , S ₂₊ , S ₃₊ , M ₁₊ , M ₃₊) Self-efficacy condom use (S ₁₊ , S ₂₊ , S ₃₊ , M ₁₊ , M ₂₊ , M ₃₊) Self-efficacy abstinence (S ₁₊ , S ₂₊ , S ₃₊ , M ₁₊ , M ₂₊ , M ₃₊) Response efficacy (S ₁₊ , S ₂₊ , S ₃₊ , M ₁₊ , M ₂₊ , M ₃₊) Response cost (S ₂₊ , S ₃₊ , M ₁₊)	No^f • No clear pattern of larger magnitude of effect, significant number, or sustained outcomes for M (vs. baseline) relative to S (vs. baseline) across outcomes and follow-up times (see previous column)
Training and wage subsidy intervention, Jordan	Groh et al., 2012 [26]	1. Job voucher (S ₁) 2. Employability training (S ₂) 3. Job voucher + employability training (M)	High quality	Follow-up at 6 and 14 mo Current employment (S ₁₊) Ever employed (S ₁₊ , S ₂₊) Months employed (S ₁₊ , S ₂₊) Hours worked last week (S ₁) Work income (S ₁) Life evaluation (current)* (S ₁₊ , M ₂₊) Life evaluation (future)* (S ₂₊) Mental health* (S ₂₊) Mobility* (S ₁₊ , M ₂₊) Empowerment* (ns) Marriage* (ns) *Examined at time 2 only	No^f • Across arm evidence (S ₁ vs. control, S ₂ vs. control, M vs. control) does not show that participation in M arm leads to a greater number or more sustained number of intended outcomes than either single-level arm (see previous column)
CASPIAN, Iran	Kargarfard et al., 2012 [33]	1. After-school physical activity program for girls (S) 2. After-school physical activity + mothers (M)	Medium quality	Follow-up at 12 wk Physiological health (resting heart rate, one-mile walk time, max. oxygen intake, flexibility, abdominal muscle strength and endurance) (S+, M+) Upper body muscle strength (S+, M+) BMI (S+, M+)	Yes^e Physiological health ↑ ^e Upper body muscle strength θ ^e BMI θ ^e • Direct statistical comparison shows M outperforms S on some physiological health indicators
School feeding program, Burkina Faso	Kazianga et al., 2009 [34]	1. School meals (S) 2. Take-home rations (conditional on attendance) (M)	High quality	Follow-up at 12 mo* New school enrollment (ns) School absenteeism (ns) Answers to math questions (ns) Time to answer math questions (ns) Cognitive development (ns) Child labor (ns) *Note: Significance reported for girls ages 13–15	No^f New school enrollment θ ^f School absenteeism θ ^f Answers to math questions θ ^f Time to answer math questions θ ^f Cognitive development θ ^f Child labor θ ^f • Comparison of S with control and M with control shows no relative difference between arms on number of intended outcomes (for girls ages 13–15 y)

(continued on next page)

Table 2
Continued

Program	Citation	Intervention arms ^b	Study quality	Significant outcomes for girls (Intervention arms vs. Baseline/control) ^c	Does multilevel arm perform better? ^d (Direct comparison of intervention arms or across arm evidence)
Entre Amigas, Nicaragua	Peña et al., 2008 [30]	1. Peer groups (S) 2. Peer groups + mothers (M1) 3. Peer groups + mothers + soap opera (M2)	Low quality	Follow-up at 15 mo Self-esteem (S+, M1+, M2+) Gender equitable attitudes (S+, M1+, M2+)	Yes^e Self-esteem θ^e Gender equitable attitudes \uparrow^e <ul style="list-style-type: none"> Girls were more likely to have positive gender visions with exposure to both peer groups and mothers (M1), regardless of soap opera component (M2) However, girl's self-esteem increased among girls in peer groups (S) relative to control, irrespective of mother's participation (M1 or M2)

^a Multilevel refers to programs that reach not just the primary target group of adolescent girls, but also include intervention activities for those who directly or indirectly affect girls' lives (e.g., parents, brothers, partners, community members).

^b Program components indicated as follows: S = single-level arm (S1 and S2 used to note more than one single-level arm, if applicable); M = multilevel arm (M1 and M2 refer to more than one multilevel arm, if applicable).

^c Outcomes: Subscripts represent multiple follow-up times (if applicable). S+/M+ = significant positive effect for girls in single-level arm or multilevel arm versus control/baseline condition; S-/M- = significant negative effect for girls in single-level arm or multilevel arm versus control/baseline condition; ns = indicates no difference for girls in any comparison group at alpha = .05 level.

^d Yes = multilevel arm outperforms single-level arm; No = multilevel arm does not outperform single-level arm. \uparrow = positive effect; θ = no difference; \downarrow = negative effect. Multilevel arm was considered to outperform a single-level arm within studies using the following levels of evidence for at least one outcome or at least one postintervention time point (if multiple).

^e Direct comparison of multilevel versus single-level arms using significance test ($p < .05$).

^f Greater number of intended significant outcomes in multilevel versus control (or baseline) than in single-level versus control (or baseline).

changes in gender attitudes. The three studies that did not find a difference were either medium- or high-quality studies. All studies that statistically compared study arms found that the multicomponent arm outperformed the single component arm [27,28,30]. Additionally, three of the studies assessed outcomes over multiple follow-up periods [26,28,29]. However, only one of these studies found that the multicomponent arm demonstrated greater effect durability for some outcomes [28].

Taken together, these data suggest that multicomponent programs outperform single component programs, although there are too few studies to conclude whether this may extend to effect durability. We note our finding is based on a small subsample of studies and half (four) were not high quality. Further, in six of the studies, the multicomponent arm (either likely or clearly) entailed more time with participating girls; of these, five [24,25,27,28,30] led to a greater likelihood of positive outcomes for girls and one did not [31]. The other two studies compared program variations that likely did not differ in the amount of time girls were exposed to the program [26,29], and in these studies, the multicomponent arm did not perform better. What degree of multicomponent program performance can be attributed to the intervention design, relative to program exposure time is thus unclear. None of the eight studies described a design that aimed to equalize the amount of time girls in different study arms were exposed to program activities.

Multilevel. Another common program design that has been used with the intent of increasing the likelihood of improved outcomes for girls is a multilevel approach—that is, programs that do not just reach adolescent girls but also include intervention activities for those who directly or indirectly affect girls' lives and well-being, such as parents, partners, husbands, and brothers. Twenty-five programs were multilevel and 28 were

single level. Five studies designed the research so that one study arm had an intervention with girls, and another study arm had that same or a similar intervention with girls plus an intervention with parents, family, or employers (Table 2) [26,30,32–34]. One, a high-quality nutrition study, found no effect among adolescent girls in our age group of interest in either study arm and we thus exclude it from further analysis [34]. Of the remaining four, two [30,33] found better outcomes in the multilevel arm—a medium-quality study based on changes in physical fitness and a low-quality study based on changes in gender attitudes. Two [26,32] (high-quality) studies found that the multilevel arm did not outperform the single-level arm. We also examined whether multilevel programs led to greater effect durability than single-level programs. Two of the studies included multiple follow-up assessments [26,32]; neither study provides evidence of more sustained multilevel versus single-level program impact.

Overall, there is insufficient evidence to conclude whether multilevel programs perform better than single-level programs. We note the limited number of studies that assess multi- and single-level program performance by design and which are of high quality (3). Notably, no studies were identified that compared an adolescent girl-only program with a variation that included the girl and her male partner.

Boosters. We defined boosters as short, supplemental activities implemented with participants sometime after the end of the main program with the aim of sustaining program effects. We only identified two studies (both medium quality and evaluating the same program) that included a booster component for participating girls [35,36]. Neither study tested the relative benefit of the booster addition, and there is insufficient evidence to determine whether this is an effective design element.

Table 3Adolescent girl programs in low- and middle-income countries that assessed program magnitude of effect by exposure level^a (N = 8)

Program	Citation	Exposure groups ^b	Study quality	Significant outcomes for girls (Intervention arms vs. Baseline/control) ^c	Is magnitude of effect greater with longer exposure? ^d (Direct comparison of exposure groups or evidence across exposure groups)
Better Life Options Program, India	Acharya et al., 2009 [41]	Intervention: – Life skills education – Livelihood training – Safe spaces girls groups • Regular attenders (half or more of sessions) (E1) • Irregular attenders (less than half of sessions) (E2) • Nonparticipants in intervention site (E3) • Control site participants (E4)	Medium quality	Independent decision-making (E1+) Self-efficacy (E1+, E4+) Mobility (E1+, E2+, E3+, E4+) Access to savings (E1+, E2+, E3+) Index of gender equitable attitudes (E1+, E2+, E3+, E4+) Gender egalitarian work attitudes (E1+, E2+) Awareness of SRH matters (E1+, E2+, E3+, E4+) Communication with parents (general topics) (E1+, E2+, E3+, E4+) Communication with parents (SRH topics) (E1+, E2+, E3+, E4+) Preferred delayed age at marriage (E1+, E2+, E3+, E4+) Mean age at marriage (E1+)	Yes ^{f,g} Independent decision-making ↑ ^{f,g} Self-efficacy ↑ ^{f,g} Mobility ↑ ^g Access to savings ↑ ^g Gender equitable attitudes ↑ ^g Gender egalitarian work attitudes ↑ ^g Awareness of SRH matters ↑ ^g Communication with parents (general topics) 0 ^g Communication with parents (SRH topics) ↑ ^g Preference for delayed marriage ↑ ^g Mean age at marriage ↑ ^f • Comparisons between 4 exposure arms at baseline vs. endline show a greater number of intended effects and/or greater magnitude of effect among group with highest program exposure (E1) • For outcomes, changes were significantly greater among E1 vs. E4 girls • Positive effect of intervention on agency and gender role attitudes was greater among E1 girls than other lower/nonexposure groups
Punjab Female Secondary School Stipend Program (FSSP), Pakistan	Alam et al., 2011 [42]	Years of CCT exposure (to girl) conditional on school attendance (E)	Medium quality	Outcomes for 3 age groups: 12–19, 15–16, 17–19 relative to control participants Middle school enrollment (E+ ages 12–19) Middle school completion (E+ ages 15–16 only) Transition to high school (E+ ages 15–16 only) Grade 9 completion (E+ ages 15–16 only) Grade 10 completion (ns) Labor force participation (E– ages 12–19 and 15–16) Work intensity (E– ages 12–19 and 15–16) Probability of delayed marriage (ns) Age of marriage (ns) Probability of birth* (ns) Number of children* (ns) *Note: Outcome reported for participants ages 17–19 only	Yes ^e Complete one grade of high school ↑ ^e Probability of delayed marriage ↑ ^e • Effects of program on education and marriage vary with length of exposure and year girls join program. Girls with 1 year or more of the program were more likely to complete 1 grade of high school and less likely to be married than girls with lower exposure levels
ISHRAQ, Egypt	Brady et al., 2007 [43]	Intervention: girl centered spaces+ literacy classes + life skills programs + sports clubs 1. Full-term participants (30 mo) (E1) 2. Dropouts (13–29 mo) (E2) 3. Dropouts (<12 mo) (E3) 4. Nonparticipants (0 mo) (E4)	Medium quality	Academic skills (writing, math, literacy) (E1+, E2+, E3+) Gender equitable attitudes on marriage (E1+, E2+) Desire for <3 children (E1+, E2+, E3+) Gender equitable attitude index (E1+, E2+) Intent to circumcise daughters (E1–, E2–, E3–) Reduced experience of FGM/C (E1+, E2+) Non-supportive attitudes toward GBV (E1+, E2+) Reduced experience of verbal abuse (E1+)	Yes ^{f,g} ↑ ^{f,g} for each outcome • Girls who dropped out early performed better than nonparticipants; full-term participants fared best of all (each group compared with baseline) both in magnitude of effect and number of significant outcomes vs. baseline across outcomes

(continued on next page)

Table 3
Continued

Program	Citation	Exposure groups ^b	Study quality	Significant outcomes for girls (Intervention arms vs. Baseline/control) ^c	Is magnitude of effect greater with longer exposure? ^d (Direct comparison of exposure groups or evidence across exposure groups)
PROGRESA/Oportunidades, Mexico	Behrman et al., 2011 [44]	1. 18+ mo of exposure to CCT (to family) conditional on girls' school attendance vs. baseline (E1) 2. <18-mo exposure to CCT (to family) conditional on girls' school attendance vs. baseline (E2)	Medium quality	Grade completion (E1+ , E2+) Employment (ns)	Yes^g Grade completion ↑ ^g Employment (0) ^g • Girls with longer exposure (vs. baseline) accumulated more schooling than those with lower exposure (vs. baseline), and this differential persisted over time (up to 5.5 y after the program start) with impacts increasing in approximately linear fashion with exposure • Greater exposure to the program had no significant effect on the proportion of girls working over time
PROGRESA/Oportunidades, Mexico	Behrman et al., 2012 [45]	1. One year exposure to CCT (to family) conditional on girls' school attendance (E1) 2. Two years exposure to CCT (to family) conditional on girls' school attendance (E2)	Medium quality	Outcomes for 3 age groups: 12–14; 15–18- and 19–20-y-old girls School enrollment (E1+ , E2+) (ages 12 to 14) Grade completion (E1+ , E2+) (all age groups) Time devoted to homework* (ns) Working for pay (E1–) (ages 15 to 18) Monthly wages (ns) *Note: Outcome not reported for 19–20-y-old age group	No^{f,g} School enrollment 0 ^g Grade completion 0 ^g Time devoted to homework 0 ^g Working for pay 0 ^f Monthly wages 0 ^f • Although there are increases in both exposure groups relative to control, there are not a greater number of intended effects or greater magnitude of effect among E2 participants (vs. control) than in E1 (vs. control)
Female Secondary School Stipend Program (FSSP), Bangladesh	Khandker et al., 2003 [46]	Continuous years of exposure (i.e., 1–4 y of implementation in school) to CCT (tuition paid to school and stipend paid to girl directly) (E) conditional on school attendance	Low quality	Secondary school enrollment (E+)	Yes^e Secondary school enrollment ↑ ^e • The program has a significant positive impact on girls' school enrollment. On average, one additional year of exposure increases girl enrollment by 8%. An additional year of future exposure increases girls' enrollment by 3%.
First Time Parents Project, India	Santhya et al., 2008 [47]	Intervention: Information provision (by outreach worker) + Social support groups + Health service adjustments 2 study sites: 1. Diamond Harbor (15% exposed to all 3 components; 51% information provision and group activities; 20% information provision only) (E1) 2. Vadodora (1% exposed to all 3 components; 9% information provision and group activities; 13% information provision only) (E2)	Medium quality	<i>Autonomy and social support</i> Role in HH decision-making (E2+) Mobility (ns) Gender role attitudes (E1+) Nonacceptability of GBV (ns) Friends in marital village (E1+) Peer support (E1+) <i>Family planning and maternal/newborn health practices</i> Index of SRH knowledge (E1+ , E2+) Married women's contraceptive use (E2+) ANC use (E1+ , E2+) Delivery preparations (E1+ , E2+) Facility-based birth (ns) PNC check within 6 wk (E1+ , E2+) Early breastfeeding adoption (E1+) <i>Partner communication and support</i> Discussed contraceptive use with partner (E1+) Partner communication in disagreement (E1+) Husband supports wife in family conflicts (ns)	Yes^f Overall ↑ ^{f,g} • Participants' endline responses were compared for those residing in each experimental site, vs. control (nonintervention) site, controlling for baseline score and other factors • Overall, a greater number of intended changes were observed in higher exposure site (E1) vs. control than lower exposure site vs. control (E2) (see previous column)

(continued on next page)

Table 3
Continued

Program	Citation	Exposure groups ^b	Study quality	Significant outcomes for girls (Intervention arms vs. Baseline/control) ^c	Is magnitude of effect greater with longer exposure? ^d (Direct comparison of exposure groups or evidence across exposure groups)
Go Girls! Initiative	Underwood and Schwandt, 2011 [48]	<u>Intervention:</u> 1. <i>Structural level</i> —training school personnel, access to financial resources to girls + families 2. <i>Community</i> —mobilization, local leadership involvement 3. <i>Family</i> —adult-child communication 4. <i>Individual</i> —community-based life skills (out-of-school girls) and school-based life skills education for boys and girls (in school) 5. Radio component (all levels) (Malawi only) Implementation areas: Botswana (E1) Malawi (E2) Mozambique (E3) Participated in at least one activity: 18% E1, 55% E2, 24% E3	Medium quality	HIV knowledge (E1+, E2+) Adult-child communication (E2+, E3+) Relationship satisfaction with mother (E1+, E2+) Reduction in teachers asking for sex in exchange for favors (E1+, E2+, E3+) Feel safe in school (ns) Legal literacy (E1+, E2+, E3+)	Yes^f Overall ↑ ^f • Overall, greatest number of intended effects among participants at endline vs. baseline in highest exposure site (E2), compared with lower exposure sites (E1 and E3), see previous column

CCT = conditional cash transfer.

^a Exposure level refers to level of participant adherence to the program, degree of program participation, or the length of time respondents receive the program.^b E refers to exposure period assessed. E1 refers to exposure group 1 in study, E2 refers to exposure group 2, etc.^c Outcomes: E+ refers to positive change relative to comparison condition (i.e., baseline or control group), whereas E− refers to negative change relative to comparison condition. ns = indicates no difference for girls in any comparison group at alpha = .05 level.^d Yes = higher exposure arm outperforms lower exposure arm; No = higher exposure arm does not outperform lower exposure arm. ↑ = positive effect; 0 = no difference; ↓ = negative effect. Higher exposure group outperforms lower exposure group using one of the following levels of evidence for at least one post-intervention time point (if multiple).^e Statistical comparison of intervention effect over time.^f Greater number of intended significant outcomes in higher exposure group versus control (or baseline) than in lower exposure group versus control (or baseline).^g Larger magnitude of effect in higher exposure group versus control (or baseline) relative to lower exposure group versus control (or baseline).

Saturation. We defined saturation as the proportion of girls in a community eligible to participate (based on program-specific criteria) who actually participated in the intervention. Ten studies included information on saturation, with reported saturation ranging from 21% [37] to close to 100% [38]. We are aware of only one study [39,40] that described varying the saturation of the program within the study design. No analysis of this variation has been published as of yet. There is thus no evidence identified in this review to inform questions of optimal program saturation for girl-centered programs. However, we note that in 9 of the 10 programs, regardless of saturation, the program had a positive effect on adolescent girls' behavior or their health or economic outcomes.

Program exposure. How much of a program's intended content is received by participants, or the program exposure level, may influence the magnitude or retention of programmatic benefit(s). Identified studies defined exposure level as participants' level of adherence to the program, degree of program participation, or, for universal programs where all eligible respondents were assumed to receive the program (e.g., a CCT intervention), the amount of time participants were exposed to the program. In total, 22 studies included information on exposure. Of these,

eight studies examining seven programs assessed the magnitude of the program effect by exposure level in the study results (Table 3). Overall, seven of the eight studies found evidence that longer exposure was associated with greater benefit [41–48]. Six of these seven were medium-quality studies, with results based on better outcomes on a range of outcome indicators, but all showed greater benefit on indicators of behavior change or impact; one was a low-quality study with results based on school enrollment. Of note is that half of identified studies were cash transfer programs measuring educational outcomes, including two studies of the same program (Progresa).¹ This limits the generalizability of the findings, although the remaining four studies reflected a greater diversity of interventions and similarly found that longer exposure tended to lead to greater improvement in outcomes at the level of the girl [41,43,47,48].

¹ The two studies of Progresa had different conclusions: Behrman (2011) found that 18 or more months of exposure to Progresa had greater effects than fewer months of exposure; and Behrman (2012) found that 1 year versus 2 years of exposure did not make a difference—both improved schooling outcomes to a similar magnitude. This may be explained by the fact that a different exposure time was contrasted, and one analysis looked at rural populations, whereas the other at urban populations.

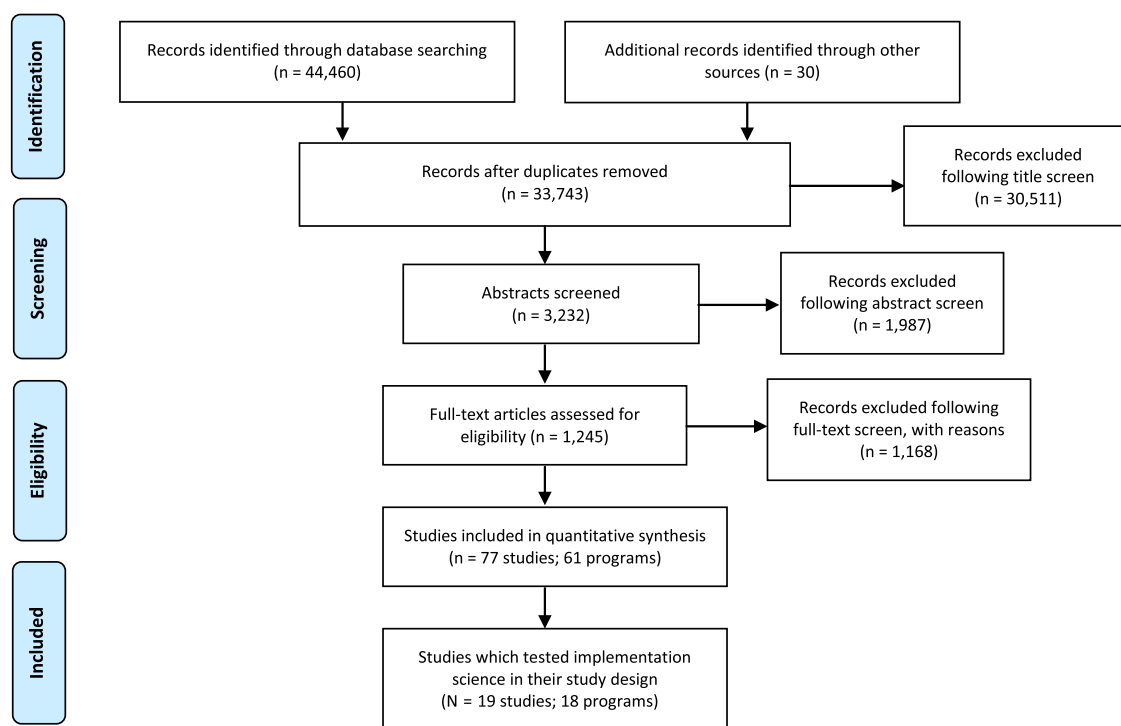


Figure 2. PRISMA flow diagram.

Two studies tested whether program results varied by exposure level using statistical tests of difference and both found evidence that longer exposure was associated with greater benefit [42,46].

Additionally, two of the included studies assessed the magnitude of program effect(s) by exposure level over multiple follow-up assessments [44,46]. Both of these studies evaluated CCT programs and found that girls with longer exposure demonstrated greater school enrollment or attainment, an effect that persisted over time. Taken together, while no studies assessing exposure were considered high quality, these data lend support to the conclusion that longer exposure matters for program effectiveness and possibly for the durability of program effects over time.

Discussion

Overall, we find that most programs for adolescent girls in LMIC demonstrated beneficial effects for girls in a variety of areas of health and well-being. We assessed the evidence on five program implementation approaches—multicomponent versus single component, multilevel versus single level, length of exposure, boosters versus no supplemental treatment, and community saturation level—to identify evidence gaps and evidence-based recommendations for the design and delivery of adolescent girls' programs on the ground. It was not possible to assess the relative influence of two strategies—boosters versus no supplemental treatment, and community saturation level—given the lack of girl-centered programs in LMIC that have tested these questions to date. Evidence was inconclusive for the question of multilevel versus single-level programs. For each program attribute considered, we did not find evidence that program

performance varied consistently by study quality. Our findings provide some evidence to suggest that multicomponent programs, relative to their single component counterparts, and longer program exposure contribute to greater intended results for girls. Although in both instances, nontrivial methodological issues temper these results. Overall, this review identifies substantial gaps in the evidence base to inform an urgent research agenda moving forward.

In five of eight studies that compared a multicomponent intervention with a single component intervention, the multicomponent intervention performed better. Considering only high-quality studies, this finding persists: three of four studies conclude that the multicomponent version worked better for girls. However, in many of these studies, it could be that the multicomponent arm achieved its effects owing to girls' longer exposure to the program (i.e., the multicomponent arm entailed more time with the girl than the single component arm). A recent study, published after our review had been completed, is notable because it addressed this challenge by equalizing the program hours of exposure across study arms [49]. Interestingly, the multicomponent arm (sexual and reproductive health plus economic program activities) did not perform definitively better than the single component arm (sexual and reproductive health only). Rather, at least in the short term, the two program variations resulted in somewhat different outcomes—smaller effects across a broader range of outcomes for the multicomponent arm and larger gains across a narrower range of outcomes for the single component arm.

The duration of program exposure also appears to matter. Seven of eight studies found that when girls were exposed to a program for a longer period of time, greater benefits accrued. However, these studies were not of high quality. For several, the

amount of exposure may reflect selection bias as contact hours were determined by the participant herself—girls who chose to attend more sessions may have been more likely to do well regardless of how many hours of intervention they received. For other programs, the exposure comparison was at the state or country level—for example, participants in one country had greater exposure than in another country—amplifying the possibility that factors unrelated to the program accounted for the observed difference in outcomes.

The evidence is extremely sparse for the booster and saturation questions. Additionally, few studies explicitly compared multilevel versus single-level programs and those that did had mixed results.

We found only two studies in LMIC that evaluated a girl-centered program with a booster, but no study that evaluated its relative efficacy. Studies in the United States have found mixed results. Some have found no additional benefit [50], but a recent, high-quality, randomized trial that used biological outcomes and included a time and dose equivalent supplement in the comparison group found that girls in the experimental condition had a significantly lower incidence of chlamydia and higher proportion of condom-protected sex acts than girls in the comparison group [51].

The question of saturation was also not tested among studies in this review. This is an area for further exploration, not just in terms of whether higher proportions of girls participating increases the likelihood of positive outcomes for participants, but also whether there is a “tipping point” at which program benefits are more likely to accrue to nonparticipants, such as younger sisters or other girls in the community.

A major limitation of this review is the methodological quality and detail of reviewed studies. A minority of studies randomized exposure to program attributes of interest, limiting our ability to infer causality. Outcomes were often self-reported by participants, which may introduce recall or social desirability bias. In general, studies provided limited information on the process of program implementation. For example, many studies did not include information on the number of hours and program length (i.e., dosage) in different study arms, or participant exposure to the intervention in practice.

There are also important limitations to our systematic review. Overall, we identified relatively few studies that addressed implementation science questions in their study design. Although some studies provided sufficient information to address implementation science questions, it was not necessarily the original objective of the authors; as a result, some of these studies yielded lower quality evidence for the purposes of our review, irrespective of the overall rigor of the study. This, together with the limited number of studies with adequate information on program design and implementation overall, resulted in this review relying on evidence that was, in some cases, based on descriptive rather than formal statistical comparisons regarding which program attribute had the strongest performance. Other aspects of implementation science such as cost and cost-effectiveness, not included in this review, are also important for program efficiency and warrant examination. Further, this review focused on girl-centered health (sexual and reproductive health, HIV, and STIs), social (including empowerment and violence), and economic programs, and is not representative of all types of interventions. Mental health and nutrition programs, for example, were not explicitly searched for. Additionally, the potential for publication bias could cause our results to be biased toward programs or out-

comes that resulted in an effect. Finally, the limited body of research identified hampered our ability to stratify results by program type or outcome.

The very good news is that this review finds that many girl-centered programs produced significant, beneficial effects for girls on a variety of health, education, and other outcomes. Girls participating in successful programs stayed in school longer, married later, had greater agency, and were more likely to find employment, among other positive effects. However, despite a relatively large initial sample of evaluations of programs for adolescent girls in LMIC, we found that few studies explicitly assessed implementation questions and methodological limitations leave many questions unanswered. As decision-makers consider girl-centered programs for replication, expansion, and scale, as well as new innovations to pilot, they need to know what works, and equally important, what does not. Filling the program implementation evidence gap is essential to guiding investment of finite resources. The cost of conducting the high-quality research needed to answer these questions is substantial, but vital to avert pouring funds into programs that do not work.

What this review presents is thus a robust and urgent research agenda to move the field of adolescent girl programming forward. Encouragingly, a few studies we identify that were published after our search suggest that implementation questions are an increasing area of inquiry. Studies need to explicitly test program permutations by design and include sufficient detail on program implementation in write-up to allow lessons to be drawn across studies. This review looked at multicomponent, multilevel, supplemental boosters, intervention exposure, and saturation. Each of these requires further rigorous research to determine whether and under what circumstances they amplify impact. There are other implementation questions such as fidelity and cost-effectiveness that are of keen interest as well. Investment in longer term follow-up would allow additional insight, including assessments of the sustainability of effects. Well-documented research from such programs is essential to maximizing resources to improve outcomes for girls.

Funding Sources

This study was part of the Population Council's RISING Program, an initiative supported by the Nike Foundation, the John D. and Catherine T. MacArthur Foundation, and the David and Lucile Packard Foundation to build the evidence base for girl-centered programs.

Supplementary Data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.jadohealth.2017.11.294>.

References

- [1] United Nations. *World population prospects: 2015 revision*. 2015. Available at: <https://esa.un.org/unpd/wpp/DataQuery/>. Accessed August 28, 2016.
- [2] Joint United Nations Programme on HIV/AIDS (UNAIDS) and The African Union. *Empower young women and adolescent girls: Fast-tracking the end of the AIDS epidemic in Africa*. Geneva: UNAIDS; 2015. Available at: http://www.unaids.org/sites/default/files/media_asset/JC2746_en.pdf. Accessed July 25, 2016.
- [3] World Health Organization. *Adolescent pregnancy factsheet*. 2014. Available at: <http://www.who.int/mediacentre/factsheets/fs364/en/>. Accessed July 10, 2016.

- [4] Mokdad AH, Forouzanfar MH, Daoud F, et al. Global burden of diseases, injuries, and risk factors for young people's health during 1990–2013: A systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2016;387:2383–401.
- [5] Steinhaus M, Fenn NS, Gregowski A, et al. "She cannot just sit around waiting to turn twenty": Understanding why child marriage persists in Kenya and Zambia. Washington (DC): International Center for Research on Women (ICRW); 2016.
- [6] UNESCO. Gender review: Creating sustainable futures for all Global Education Monitoring Report. Paris: UNESCO; 2016.
- [7] World Health Organization. Global and regional estimates of violence against women. Geneva: WHO; 2013.
- [8] McCarthy KJ, Hallman K, Brady M. Investing when it counts: Reviewing the evidence and charting a course of research and action for very young adolescents. New York: Population Council; 2016.
- [9] Sawyer SM, Afifi RA, Bearinger LH, et al. Adolescence: A foundation for future health. *Lancet* 2012;379:1630–40. doi:10.1016/S0140-6736(12)60072-5.
- [10] Culter DM, Llears-Muney A. Education and health: Evaluating theories and evidence. NBER Working Paper No. 12352. 2006. doi:10.3386/w12352.
- [11] Campbell JC. Health consequences of intimate partner violence. *Lancet* 2002;359:1331–6. doi:10.1016/S0140-6736(02)08336-8.
- [12] Nour N. Health consequences of child marriage in Africa. *Emerg Infect Dis* 2006;12:1644–9. doi:10.3201/eid1211.060510.
- [13] Levin R, Lloyd CB, Greene M, et al. Girls count: A global investment and action agenda. Washington (DC): Center for Global Development; 2009.
- [14] Temin M, Levin R. Start with a girl: A new agenda for global health. Washington (DC): Center for Global Development; 2009.
- [15] Hallman K, Stoner M, Chau M, Melnikas AJ. A review of control-comparison interventions on girls and health in low and middle-income countries. London, UK: Girl Effect; 2013.
- [16] Hardee K, Gay J, Croce-Galis M, et al. What HIV programs work for adolescent girls? *J Acquir Immune Defic Syndr* 2014;66:S176–85. doi:10.1097/QAI.0000000000000182.
- [17] Hennegan J, Montgomery M. Do menstrual hygiene management interventions improve education and psychosocial outcomes for women and girls in low and middle income countries? A systematic review. *PLoS ONE* 2016;66:e0146985. doi:10.1371/journal.pone.0146985.
- [18] Kalamar AM, Lee-Rife S, Hindin MJ. Interventions to prevent child marriage among young people in low- and middle-income countries: A systematic review of the Published and Gray Literature. *J Adol Health* 2016;59:S16–21. doi:10.1016/j.jadohealth.2016.06.015.
- [19] Hindin MJ, Kalamar AM, Thompson TA, et al. Interventions to prevent unintended and repeat pregnancy among young people in low- and middle-income countries: A systematic review of the published and gray literature. *J Adol Health* 2016;59:S8–15. doi:10.1016/j.jadohealth.2016.04.021.
- [20] Lundgren R, Amin A. Addressing intimate partner violence and sexual violence among adolescents: Emerging evidence of effectiveness. *J Adol Health* 2015;56:S42–50. doi:10.1016/j.jadohealth.2014.08.012.
- [21] Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ* 2009;339:b2535. doi:10.1136/bmj.b2535.
- [22] What works for women and girls. 2015. Available at: <http://www.whatworksforwomen.org/pages/methodology>. Accessed June 24, 2015.
- [23] Effective Public Health Practice Project (EPHPP). Quality rating tool. 2015. Available at: <http://www.ephpp.ca/tools.html>. Accessed June 24, 2015.
- [24] Amin S, Suran S. Program efforts to delay marriage through improved opportunities: Some evidence from rural Bangladesh. New York: Population Council; 2005.
- [25] Austrian K, Muthengi E. Safe and smart savings products for vulnerable adolescent girls in Kenya and Uganda: Evaluation report. Nairobi: Population Council; 2013.
- [26] Groh M, Krishnan N, McKenzie D, Vishwanath T. Soft skills or hard cash? The impact of training and wage subsidy programs on female youth employment in Jordan. Policy Research Working Paper 6141. Washington (DC): World Bank Group; 2012. doi:10.1596/1813-9450-6141.
- [27] Hallman K, Roca E, Siyakha Nentsha: Building economic, health, and social capabilities among highly vulnerable adolescents in KwaZulu-Natal, South Africa. New York: Population Council; 2011.
- [28] Hallfors D, Cho H, Rusakaniko S, et al. Supporting adolescent orphan girls to stay in school as HIV risk prevention: Evidence from a randomized controlled trial in Zimbabwe. *Am J Public Health* 2011;101:1082–8. doi:10.2105/AJPH.2010.300042.
- [29] Montgomery P, Ryus CR, Dolan CS, et al. Sanitary pad interventions for girls' education in Ghana: A pilot study. *PLoS ONE* 2012;7:e48274. doi:10.1371/journal.pone.0048274.
- [30] Peña R, Quintanilla M, Navarro K, et al. Evaluating a peer intervention strategy for the promotion of sexual health-related knowledge and skills in 10- to 14-year-old girls. Findings from the "Entre Amigas" project in Nicaragua. *Am J Health Promot* 2008;22:275–81. doi:10.4278/06012310R3.1.
- [31] Scales PC, Benson PL, Dershem L, et al. Building developmental assets to empower adolescent girls in rural Bangladesh: Evaluation of project Kishoree Kontha. *J Res Adolesc* 2013;23:171–84. doi:10.1111/j.1532-7795.2012.00805x.
- [32] Pham V, Nguyen H, Tho LH, et al. Evaluation of three adolescent sexual health programs in Ha Noi and Khanh Hoa Province, Vietnam. *AIDS Res Treat* 2012;2012:1–12. doi:10.1155/2012/986978.
- [33] Kargarfard M, Kelishadi R, Ziaee V, et al. The impact of an after-school physical activity program on health-related fitness of mother/daughter pairs: CASPIAN study. *Prev Med* 2012;54:219–23. doi:10.1016/j.ypmed.2012.01.010.
- [34] Kazianga H, de Walque D, Alderman H. Educational and health impacts of two school feeding schemes: Evidence from a randomized trial in rural Burkina Faso. Policy Research Working Paper 4976. Washington (DC): World Bank Group; 2009. doi:10.1596/1813-9450-4976.
- [35] Sarnquist C, Omondi B, Sinclair J, et al. Rape prevention through empowerment of adolescent girls. *Pediatrics* 2014;133:e1226–32. doi:10.1542/peds.2013-3414.
- [36] Sinclair J, Sinclair L, Otieno E, et al. A self-defense program reduces the incidence of sexual assault in Kenyan adolescent girls. *J Adolesc Health* 2013;53:374–80. doi:10.1016/j.jadohealth.2013.04.008.
- [37] Bandiera O, Buehren N, Burgess R, et al. Empowering adolescent girls: Evidence from a randomized control trial in Uganda. London: London School of Economics and Political Science; 2012.
- [38] Mulralidharan K, Prakash N. Cycling to school: Increasing secondary school enrollment for girls in India. Discussion Paper No. 7585. Bonn: Institute for Study of Labor (IZA); 2013.
- [39] Baird S, McIntosh C, Ozler B. Cash or condition? Evidence from a cash transfer experiment. *Q J Econ* 2011;126:1709–53. doi:10.1093/qje/qjr032.
- [40] Baird SJ, Garfein RS, McIntosh CT, et al. Effect of a cash transfer programme for schooling on prevalence of HIV and herpes simplex type 2 in Malawi: A cluster randomised trial. *Lancet* 2012;379:1320–9. doi:10.1016/S0140-6736(11)61709.
- [41] Acharya R, Kalyanwala S, Jejeebhoy S, et al. Broadening girls' horizons: Effects of a life skills education programme in rural Uttar Pradesh. New Delhi: Population Council; 2009.
- [42] Alam A, Baez JE, Del Carpio XV. Does cash for school influence young women's behavior in the longer term? Evidence from Pakistan. Policy Research Paper 5669. Washington (DC): World Bank Group; 2011. doi:10.1596/1813-9450-5669.
- [43] Brady M, Assad R, Ibrahim B, et al. Providing new opportunities to adolescent girls in socially conservative settings: The Ishraq program in rural Upper Egypt. New York: Population Council; 2007.
- [44] Berhman JR, Parker SW, Todd PE. Do conditional cash transfers for schooling generate lasting benefits?: A five-year follow-up of PROGRESA/Oportunidades. *Journal of Human Resources* 2011;46:203–36. doi:10.1353/jhr.2011.0028.
- [45] Behrman JR, Gallardo-Garcia J, Parker SW, et al. Are conditional cash transfers effective in urban areas? Evidence from Mexico. *Educ Econ* 2012;20:233–59. doi:10.1080/09645292.2012.672792.
- [46] Khandker S, Pitt M, Nobuhiko F. Subsidy to promote girls' secondary education: The female stipend program in Bangladesh. Munich Personal RePEc Archive No. 23688. Munich: Munich University Library; 2003.
- [47] Santhya KG, Haberland N, Das A, et al. Empowering married young women and improving their sexual and reproductive health: Effects of the First-time Parents Project. New Delhi: Population Council; 2008.
- [48] Underwood C, Schwandt H. Go girls! Initiative research findings report: Data from the 2009 baseline survey, 2009–10 process evaluation and 2010 endline survey in Botswana, Malawi and Mozambique. Baltimore: John Hopkins Bloomberg School of Public Health/Center for Communication Programs; 2011.
- [49] Edmeades J, Lantos H, Mekuria F. Worth the effort? Combining sexual and reproductive health and economic empowerment programming for married adolescent girls in Amhara, Ethiopia. *Vuln Child Youth* 2016;11. doi:10.1080/17450128.2016.1226529.
- [50] Foshee V, Bauman KE, Ennett ST, et al. Assessing the long-term effects of the Safe Dates Program and a booster in preventing and reducing adolescent dating violence victimization and perpetration. *Am J Public Health* 2004;94:619–24.
- [51] DiClemente RJ, Wingood GM, Sales JM, et al. Efficacy of a telephone-delivered sexually transmitted infection/human immunodeficiency virus prevention maintenance intervention for adolescents. *JAMA Pediatr* 2014;168:938–46. doi:10.1001/jamapediatrics.2014.1436.