IDS Bulletin
Transforming Development Knowledge
Volume 49 | Number 4 | September 2018

THE MILLENNIUM VILLAGES: LESSONS ON EVALUATING INTEGRATED RURAL DEVELOPMENT

Editor **Chris Barnett**

Applying Factorial Designs to Disentangle the Effects of Integrated Development*†

Holly M. Burke,[1] Mario Chen[2] and Annette N. Brown[3]

Abstract In this article, we discuss the study design and lessons learned from a full-factorial randomised controlled study conducted with beneficiaries of a youth programme in Pretoria, South Africa. The study assesses whether the integration of an economic strengthening intervention with an HIV-prevention education intervention improves economic and health outcomes beyond singular interventions. The selected youth were randomised into four groups: combined economic strengthening and HIV-prevention interventions; economic strengthening intervention only; HIV-prevention education intervention only; or no interventions. We conducted a pre-intervention and two post-intervention assessments with the participants to measure outcomes, including the primary outcome – prevalence of sexually transmitted infections. We discuss our rationale for the study design and the challenges faced when implementing it. We consider how features of the integrated programme, such as how synergy is assessed, and features of context, for example available sample size, determine which methods can be used to test the effectiveness of integrated programming.

Keywords: integrated; development; multidisciplinary; multisector; evaluation; synergy; interaction effects; HIV prevention; economic strengthening.

1 Background

Globally, an estimated one third of all new HIV infections occurs among youth aged 15–24, highlighting the importance of an HIV response targeting youth (UNICEF 2013). Evidence shows that girls who engage in intergenerational and transactional sex are especially vulnerable to HIV (Luke 2005; Leclerc-Madlala 2008). Several studies show that HIV prevention education can educate and build skills, which leads to safer sex practices and lower rates of HIV and other sexually transmitted infections (STIs) (Jewkes *et al*. 2006; Kirby, Laris and Rolleri 2007; Wingood *et al.* 2007). There is also evidence that

The *IDS Bulletin* is published by Institute of Development Studies, Library Road, Brighton BN1 9RE, UK
This article is part of *IDS Bulletin* Vol. 49 No. 4 September 2018 **'The Millennium Villages: Lessons on Evaluating Integrated Rural Development'**; the Introduction is also recommended reading.

economic strengthening (ES) interventions can foster greater financial independence, which may reduce the incentive to participate in transactional and intergenerational sex, and increase bargaining power in sexual relationships, for example by insisting on condoms (Swann 2018). Swann (2018) reviews a large body of evidence and concludes that cash transfers and educational support are effective in reducing self-reported HIV risk behaviours, especially among adolescents.

However, clinical evidence supporting these behavioural findings is currently lacking, and evidence for other types of economic strengthening interventions that may be more sustainable, such as savings and financial education, is even less conclusive (*ibid*.). Furthermore, HIV prevention and economic strengthening interventions are often conducted in isolation, despite HIV having both economic and social drivers. Some research has yielded promising results suggesting that interventions with economic and social components build skills to improve financial wellbeing, women's empowerment, and reduce intimate partner violence, thus reducing the vulnerabilities of HIV (Pronyk *et al*. 2006; Kim *et al*. 2007; Gupta *et al*. 2013). Kim *et al.* (2009) conducted a cluster randomised trial of a gender and HIV training programme combined with a microfinance intervention implemented in rural South African villages, and found greater reductions in women's HIV risk behaviours compared to groups receiving only the microfinance intervention and no intervention. However, few studies have investigated whether there is synergy in multisectoral integration; that is, whether the integration of two interventions produces a combined effect greater than the sum of the interventions' separate effects (Ahner-McHaffie *et al*. 2018).

To build the evidence base around the efficacy of integrated interventions for HIV prevention, we conducted a full factorial randomised controlled study to assess whether the integration of an economic strengthening intervention with an HIV prevention education intervention improves health and economic outcomes for adolescents beyond singular interventions. In this article, we discuss our rationale for the study design and the challenges we faced when implementing it. We describe our analysis plan as designed and how it can and cannot be adjusted to account for the implementation challenges. We discuss how features of the integrated programme, such as how synergy is assessed, and how features of the context, such as the available sample size, determine which methods can be used to test the effectiveness of an integrated model within a full factorial design.

2 Study overview

2.1 Study participants

The study was conducted with the adolescent beneficiaries of a local, not-for-profit organisation implementing a programme providing support to poor families affected by HIV in four communities in Gauteng Province, South Africa.

2.2 Intervention description

FHI 360's Accelerating Strategies for Practical Innovation and Research in Economic Strengthening (ASPIRES)[4] project provided technical assistance to the programme implementer to implement the economic strengthening and HIV prevention interventions, which were both educational interventions. The economic strengthening intervention – Impumelelo – builds on the Life Poa curriculum developed by YouthSave Kenya, and covered the topics of budgeting and saving, education on different savings options, and introduced the topic of earning money. The HIV prevention intervention, an updated version of the existing Vhutshilo curriculum that has been used with vulnerable youth throughout South Africa, covered expressing feelings, dealing with loss and grief, decision-making, drugs and alcohol, HIV and other STIs, healthy relationships, safer sex, and contraception. Each intervention consisted of 16 in-person sessions with a group of approximately 15 youth that lasted approximately 90 minutes. Participants in the combined intervention group received the two interventions sequentially (32 sessions total), though the order of the interventions varied due to programme staffing constraints.

2.3 Study design

From January to July 2016, 1,773 females and males aged 14–17 years were randomised in a 1:1:1:1 ratio to the four study groups: ES and HIV (combined economic strengthening and HIV prevention education interventions), ES (economic strengthening intervention only), HIV (HIV prevention education intervention only), or control (no additional education interventions). All participants received the programme's standard package of services which includes food and education support and linking families with community services, such as access to government grants. This trial is registered with ClinicalTrials.gov, number NCT 02888678.

2.4 Data collection

Employing a panel design, we collected data from the same participants three times during the study: at baseline (before the interventions began) and at two endlines. Endline 1 occurred shortly after the interventions finished and endline 2 occurred approximately eight months later to investigate the sustainability of the treatment effect. During each round of data collection, participants completed an in-person behavioural survey using audio computer-assisted self-interview (ACASI) technology where participants can read questions on a computer screen, hear the questions read to them privately through headphones, and indicate their responses without anyone viewing their selections during the interview. Participants also provided a urine sample for STI and pregnancy testing. Participants with positive STI or pregnancy tests were notified and referred for treatment or services after each round of data collection. Data collection ended May 2018 and analysis is ongoing.

We documented the implementation processes and costs associated with implementing the interventions. Programme staff, with support

from research staff, documented the resources used to implement the interventions at the programme level using electronic spreadsheets specifically tailored for the study. We also interviewed 22 purposively selected programme staff to gather information about the inputs needed to implement the interventions and their perspectives on the implementation of the interventions, including challenges experienced and resolutions to those challenges. Interviewees included the project coordinator, master trainer, programme managers, and facilitators of the economic strengthening intervention and the HIV prevention intervention.

2.5 Outcomes

The primary outcome for testing the effectiveness of the integration model is STI prevalence, defined as a positive test result for gonorrhoea, trichomoniasis, or chlamydia infection. We selected these non-viral STIs because they are common and treatable, and can be tested in urine samples obtained from both males and females. The secondary outcomes (all self-reported, except pregnancy) are:

1 Pregnancy (performed on the urine specimens of female participants);

2 Engaging in protective sexual behaviour, defined as self-reported abstinence or consistent condom use over the past six months;

3 Engaging in transactional sex in the past six months;

4 Having two or more sexual partners in the past six months;

5 HIV knowledge;

6 Financial literacy;

7 Participation in a savings group;

8 Opening a savings account;

9 Net change in savings in past year;

10 Saving for education;

11 Caregiver being primary provider of money to youth for savings;

12 Participation in household budgeting.

3 Challenges and approaches to evaluating an integrated programme

Evaluations of complex programmes, such as those with integrated multisectoral interventions, face important challenges. In this section, we discuss the challenges we faced and the decisions we made when designing and implementing the evaluation of a programme integrating interventions from the health and economic sectors. We specifically discuss how synergy is assessed, sample size considerations, and our analysis strategy and outcomes.

3.1 Challenge 1: how to define and measure integration and synergy effects

To assess the effect of an integrated programme against single interventions, we need to compare the integrated programme with each intervention implemented separately. A control group with no intervention or standard of care is also needed to determine the net effect of each single intervention and to determine how much better (or worse) the integrated programme is in affecting the outcomes. A systematic review found, however, that most experimental evaluations of integrated development programmes are 'two-arm' studies, comparing a group participating in the integrated programme to a control group not participating in the programme (FHI 360 2014). The review concluded that these comparisons preclude any assessment of whether single interventions achieve similar results as the integrated programme or what effects are attributable specifically to the integration.

A second systematic review, focused more specifically on this concern, assessed whether studies evaluating integrated development programmes measured synergistic effects (Ahner-McHaffie *et al.* 2018). Two programmes are said to work synergistically if the effects of the integrated intervention are amplified beyond the sum of the effects of each single sector intervention. Among the 601 impact evaluations included in this second review, 12 used partial factorial designs, and 26 used full factorial designs. In a full factorial design (a '2x2 design', assuming the integrated programme combines two interventions), the evaluation analyses data across four arms (or participants' groups), including separate arms for each of the interventions alone, for the programme that integrates those interventions, as well as for a control group. Only those impact evaluations with full factorial designs allow the measurement of the impact from integration and from synergy. The review finds, however, that most of the full factorial studies do not clearly discuss the distinct effects of synergy.

In our study, we used a full 2x2 factorial design and randomly assigned participants to ES, HIV, ES + HIV, or control. The synergy question is whether the whole is greater than the sum of the parts and can be stated as: is $1 + 1 > 2$? In a straightforward linear model, we can estimate the effect of each intervention, as well as the effect of the integrated approach using interaction terms to assess the effects of different interventions, whether implemented singly or in combination. We can consider the different effect scenarios as:

$1 + 1 = 2$, there is no synergy effect;

$1 + 1 > 2$, there is synergy (amplifying effects);

$1 + 1 < 2$, there is a detrimental effect.

However, even in the presence of detrimental effects (i.e. integrated programme not achieving the full sum of the single intervention effects), the integrated programme may still be considered beneficial if it improves outcomes more than each of the single interventions.

Put simply, it may still be the case that the whole is greater than either of the parts separately. Therefore, we posit another effect of interest that is less stringent for determining the value of the integrated programme: $1 + 1 > 1$. That is, we posit that integration produces a positive effect on top of the single intervention effect, even if the integration does not produce synergistic effects. This equation answers the question of whether the integrated ES and HIV intervention improves outcomes beyond what could be achieved by implementing either of the single interventions alone. It tells us if the integrated programme is the most effective of the three possibilities.

For our study, we based the sample size calculations on detecting $1 + 1 > 1$ to focus on the effects of the integrated programme and to mitigate the demand on sample size for adequately assessing the synergy hypothesis (see Section 3.2). To be exact, we focused on testing the two-sided version $1 + 1 \neq 1$ to allow for the possibility that integrating the programmes undermines the effect that could be achieved if we keep the interventions separated. This undermining effect of the integration could happen, for example, if programme staff or youth become overwhelmed by having too much to do in the integrated programme and therefore underperform in both components.

3.2 Challenge 2: the need for a large sample

Integrated development evaluations using a factorial design create several challenges for sample size. The first is that multi-arm studies divide the total sample into more groups than a two-arm (programme and control) study. As noted in Section 2.3, we divided our sample of 1,773 participants into four groups. If we think of simply testing each treatment arm against the control, we are only using half of the total sample for each test. We would need a factorial design sample size twice as large as the two-arm study to get the same power to measure the effect of the integrated programme against the control.

Another challenge comes from the potential of an interaction effect between the two individual interventions, where a positive interaction effect indicates synergy, one of the hypotheses we would like to test whenever possible. The challenge comes from the need for a larger sample size to detect interaction effects. Wolbers *et al.* (2011) provide sample size requirements for different levels of interaction effects. For example, they found that even under large interaction effects, doubling (strong synergy) or nullifying (zero effect of the integration) the effects of the single interventions requires fourfold the sample size of a two-group study (*ibid.*).

Unfortunately, we were not able to draw a sample large enough to test – with sufficient power – for interaction effects, at least not based on the assumptions in our power calculations. To address this, our primary hypothesis testing strategy will simply compare the integrated programme to each of the interventions separately. Thus, to conclude that the ES + HIV programme is effective, we will test

whether the integrated programme is statistically significantly more effective than each one of the interventions implemented separately. That is, ES + HIV compared to ES and ES + HIV compared to HIV. A statistically significant positive result in both comparisons will indicate a benefit associated with the integration model over what could be achieved with either of the interventions implemented separately. A statistically significant negative result in both tests will indicate a harmful effect of the integration. Effects in different directions may also indicate integration failure. We will also use the full data set to test whether there is a synergistic effect of the integrated programme (i.e. positive interaction effect), but understanding that we are likely to be underpowered to detect this effect.

Another challenge to sample size for evaluating integrated development programmes is multiple outcome testing, also called multiple comparisons or multiple inferences. The factorial design introduces multiple comparisons just based on the design alone. But even without a factorial design, evaluations of integrated programmes are likely to include measurements of many outcomes. Multiple outcome measurements arise from the desire to assess outcomes directly related to individual interventions and possible additional outcomes from the integrated programme. In Section 2.5, we present one primary and 12 secondary outcomes for our study.

The challenge arises because the more outcomes you test with the same data, the more likely you are to find statistically significant results for one or a few outcomes by chance alone. In statistical terms, the multiple comparisons problem leads to type I error inflation; you are more likely to reject a null hypothesis that is true (find an effect that is not there). The solutions to this problem require a more complex analytical model. For the more complex model to have the same power as a single comparison study, you need a larger sample size.

Based on the sample size we have, our strategy for addressing the multiple comparisons challenge is to pre-specify a primary outcome that we will use for our causal inferences. As noted above, this outcome is STI prevalence at each endline (with all positives at baseline and endline 1 receiving treatment so that prevalence starts at zero). We will present the analysis of the other 12 outcome variables as exploratory.

One more design consideration affecting sample size and statistical power for integrated development evaluations is that complex programmes are often implemented in groups, such as schools or clinics. If the randomisation is at the group or cluster level, then a larger sample is needed. Clustered designs are statistically less efficient because units within clusters are expected to be more homogeneous than they would be across clusters.

We were fortunate in our study to avoid a cluster randomised design. For our interventions, individual randomisation makes sense. The

interventions are designed to change individuals' behaviours, as opposed to programmes designed to change outcomes at a group level (for example, improving school quality or agricultural markets). The intervention activities (sessions), however, were designed to be carried out in groups. The programme implementer had no pre-existing groups that could be used for delivering the study interventions, so these needed to be formed for our study. Working closely with the programme implementer, we were able to recruit participants into our study, and randomise them into study arms, before they were then put into groups formed within each arm to receive the interventions (or not, in the case of the control arm). However, we recognise that because the intervention was delivered at the group level, group level differences may arise (for example, if some programme staff are better at delivering the interventions than others). If these group effects exist, we will control for them in our model and thus may lose some power for detecting the effects of interest. It is important to note that if we do find that the impact of the integrated programme is highly dependent on the facilitators' performance, it will imply that the programme is less scalable and potentially less useful for preventing HIV.

3.3 Challenge 3: implementation

A 2x2 factorial design requires three different implementations plus additional recruitment and data collection for the control. That means that the typical challenges studies face due to implementation are multiplied. Our study provides several examples. First, the need for the largest possible sample caused our timeline to be delayed because of the time it took to enrol nearly 1,800 eligible youth. We also faced delays from the need to hire the requisite staff to implement the two completely different and time-intensive education interventions. The implementer had to run three programme cycles to serve all the youth assigned to the intervention arms. The need for this repetition was driven in large part by the ES and HIV group, which required 32 separate training sessions.

Second, instead of the typical research format where the baseline and endline assessments are conducted for all participants at the same time and the interventions are implemented in between, we needed to enrol participants and collect the data on a continuous basis to complete the study on time. In addition to increasing the research costs, this meant that the first half of participants started interventions while we were still enrolling the second half of participants into the study. Moreover, at the end of the second intervention cycle, we began data collection for the first endline with the first group of enrolled participants (since they finished their interventions). This meant that some of the participants waited months between enrolment and their intervention to begin, while others started their intervention right after enrolment. These delays in a situation of multiple education interventions could result in contamination across intervention arms, as participants who know they will be taking a course but have to wait may seek discussions with others who are already in courses, regardless of whether it is the same course (i.e. study arm). Multiple intervention cycles also meant that

implementation could change over time, for example, as programme staff become more familiar with the curricula and become more experienced facilitators.

The third challenge to our study design from implementation constraints is that the implementer used a separate set of staff members to deliver the economic strengthening intervention from those who delivered the HIV prevention intervention. This specialisation facilitated the management of the work load and increased the quality and uniformity of delivery of the interventions. The HIV prevention intervention requires facilitators with higher skill levels, however, because they need to facilitate group sessions on the topics of HIV and sexual behaviour, which are more sensitive and stigmatised than topics like financial literacy and savings. In our 2x2 design then, the comparison of the two standalone intervention arms could be confounded or moderated by the quality of the facilitators.

3.4 Challenge 4: measuring outcomes

To fully evaluate this integrated programme, we needed to measure both economic and health outcomes. This meant we needed to collect a lot of data because these two sectors use very different indicators, methods, and timelines to measure outcomes. We settled on 13 outcome indicators, six health indicators, and seven economic indicators. To address multiple comparison concerns as mentioned above, we chose a primary outcome, STI prevalence, because it is a marker of unprotected sex, which is also the main risk factor for HIV transmission in the study setting. This clinical outcome is also considered less biased than self-reported measures and can be reliably measured in both boys and girls.

ES and HIV risk behaviours require different measurement techniques. Sexual behaviour that puts people at risk of acquiring HIV, such as engagement in transactional sex, is challenging to accurately measure through self-report because it is stigmatising, and sex work is illegal in many contexts. This required us to utilise additional (and often costly) technology to reduce reporting bias. In our study, we used ACASI techniques coupled with testing biological specimens for STIs and pregnancy. Economic strengthening outcomes, on the other hand, are less stigmatised and therefore may be more readily obtained through self-report.

ES and HIV risk behaviours may also develop differently over time, and this required us to take more than one endline measurement. For example, after participating in the ES intervention, youth may start to save money; however, it will take most youth a long time to save enough money to obtain higher education or skills training, start a business, or acquire enough productive assets to become financially independent to the point that they no longer need to engage in transactional or intergenerational sex to meet their needs. Whereas after participating in the HIV prevention intervention, youth may be more likely to engage in protective behaviours such as using condoms, with prevention messages

fresh in their minds. As time passes from the last lesson, we expect that sexual risk-taking will increase as the messages are forgotten or as other needs and priorities become more immediate, and as the youth mature into young adults. To investigate these varying timelines, we collected endline data twice: once right after the intervention ended and again as far out as our grant allowed. Three data points will also allow us to explore trends and the sustainability of effects over time.

3.5 Challenge 5: adherence and loss to follow-up

Adherence is a challenge for most interventions, not least of which are interventions that involve youth attending multiple group sessions after school. The challenge is greater for integrated programmes like ours where the integration is additive, because the integrated group has more to adhere to compared to the single intervention groups. It is also possible that one intervention type may have higher adherence than another because it has fewer requirements or is more desirable to the target population.

While participant retention is critical in the evaluation of any intervention, a time-intensive integrated intervention, like the one we evaluated, has the potential to result in differential loss to follow-up if participants in the integrated group are more likely to drop out of the study compared to other groups. Fortunately, through the diligent work of our research staff and support from the programme implementer, we had high overall participant retention throughout our study: 88 per cent at the first endline and 86 per cent at the second endline. We have not examined retention for each of our groups as of writing this article, but we are not expecting differential loss to follow-up, given our high overall retention. Low, non-differential loss to follow-up will reduce bias in our findings and gives us the best chance of reliably testing our research hypotheses.

4 Discussion

Integrated programmes that include economic strengthening components are increasingly being implemented to prevent HIV in resource-limited settings, but without rigorous evidence supporting this approach. We implemented a full factorial randomised controlled study to build the evidence base around the efficacy of integrated programmes for HIV prevention. While our study, like most evaluations conducted in development settings, faced financial and logistical constraints that prevented us from gathering a larger sample, we still believe the factorial design was the right decision. Unfortunately, we have not found any factorial design studies of integrated programmes that were able to draw sufficient samples to fully address all the challenges, but we hope that more integrated development programmes will be evaluated using factorial designs in the future. Given the statistical challenges, however, the results from these studies should be carefully interpreted. P-values can be easily misinterpreted if they are not clearly linked to the specific effects that are associated with hypotheses of interest in the evaluation of integrated programmes.

In designing this 2x2 factorial study of an integrated development programme, we felt that assessing synergy, sample size, outcome measurement, adherence, and loss to follow-up would be our main challenges. As discussed, we were not able to fully resolve the sample size challenges, which arise from multiple arms and multiple outcomes. Also, testing for synergy requires larger sample sizes. Our study protocol describes our analytical approaches for addressing these challenges in more detail. One approach we recommend is to focus the evaluation of integrated programmes on determining whether integration provides benefits beyond single interventions even if not amplifying the effects as expected under synergistic effects. Although ideal, synergy (i.e. the whole being greater than the sum of its parts) should not necessarily be the goal of every integrated programme. In terms of outcome measurement, we prioritised, at great expense, extensive data collection, including clinical indicators for STIs. Through careful implementation, we avoided major problems from loss to follow-up. What we did not fully anticipate were challenges to the study from the implementation constraints, such as differences among facilitators, for a multi-intervention, complex programme. In our analysis, we will pay particular attention to whether any of those challenges ultimately bias or contaminate our results.

Our main recommendation for future evaluations employing the full factorial design is to carefully consider all possible options to maximise the sample size. One approach to free up resources for a larger sample size is to not collect baseline data, as randomised controlled trials do not require baseline measures to help control for bias. That would not have worked in our situation, as we needed to treat all the baseline STIs to be able to measure STI prevalence as our outcome for endlines 1 and 2. It could work in other situations, though. Also, depending on the outcome targeted, researchers might consider collecting data for only one endline. In our case, we had two very different interventions that are on different timelines, so we saw value in collecting data for more than one endline, but programmes that integrate interventions with similar theories of change or similar timelines, especially interventions from the same sector, may not have this need.

While we were able to collect extensive data on a variety of outcomes for our full sample to conduct quantitative analysis, we recognise that mixed methods designs allow better analysis of the hows and whys, particularly for understanding synergistic effects. Unfortunately, funding cuts and unanticipated, increased expenses due to timeline delays forced us to eliminate our qualitative research component from the evaluation. Our plan was to conduct in-depth interviews with programme staff, youth participants, and their caregivers to describe whether, how, and why the interventions were perceived as effective in addressing economic and health outcomes among youth. Despite losing this important research component, the structured interviews we conducted with staff about their experience implementing the interventions gave us insight into how the interventions worked from the perspective of the implementers, and will contribute to the interpretation of the final

results. However, our recommendation is to include robust qualitative components in future integrated development evaluations.

We hope that the discussion of the challenges we encountered and our approaches for mitigating them contributes to the ongoing discussion on how integrated programmes should be evaluated and findings interpreted. Framing the interpretation of the results on the estimable effects in the context of study limitations will be important for the appropriate use of the findings.

Notes

* This issue of the *IDS Bulletin* was prepared as part of the impact evaluation of the Millennium Villages Project in northern Ghana, 2012–17, funded by the UK Department for International Development (DFID) (www.dfid.gov.uk). The evaluation was carried out by Itad (www.itad.com) in partnership with IDS (www.ids.ac.uk) and PDA-Ghana (www.pdaghana.com). The contents are the responsibility of the evaluation team and named authors, and do not necessarily reflect the views of DFID or the UK Government.

† The authors wish to acknowledge the following FHI 360 colleagues for their helpful review of this commentary: Mandy Swann, Emily Namey, and Michael Ferguson.

1 Corresponding author. FHI 360, Durham, North Carolina, USA. 359 Blackwell Street, Suite 200, Durham NC 27701, USA. hburke@fhi360.org.

2 FHI 360, Durham, North Carolina, USA. 359 Blackwell Street, Suite 200, Durham NC 27701, USA. mchen@fhi360.org.

3 FHI 360, Washington, District of Columbia, USA. 1825 Connecticut Avenue, NW, Suite 8000, Washington DC 20009, USA. abrown@fhi360.org.

4 The Accelerating Strategies for Practical Innovation and Research in Economic Strengthening (ASPIRES) project, supported by the President's Emergency Plan for AIDS Relief (PEPFAR) and the United States Agency for International Development (USAID) and managed by FHI 360, supports gender-sensitive programming, research, and learning to improve the economic security of highly vulnerable individuals, families, and children.

References

Ahner-McHaffie, T.W.; Guest, G.; Petruney, T.; Eterno, A. and Dooley, B. (2018) 'Evaluating the Impact of Integrated Development: Are We Asking the Right Questions? A Systematic Review', *Gates Open Research* 1.6: 1–23, http://dx.doi.org/10.12688/gatesopenres.12755.2 (accessed 20 September 2018)

FHI 360 (2014) *Integration of Global Health and Other Development Sectors: A Review of the Evidence*, www.fhi360.org/resource/integration-global-health-and-other-development-sectors-review-evidence-summary-brief-full (accessed 13 September 2018)

Gupta, J. *et al.* (2013) 'Gender Norms and Economic Empowerment Intervention to Reduce Intimate Partner Violence Against Women

in Rural Côte d'Ivoire: A Randomized Controlled Pilot Study', *BMC International Health and Human Rights* 13.46: 1–12

Jewkes, R. *et al.* (2006) 'A Cluster Randomized-Controlled Trial to Determine the Effectiveness of Stepping Stones in Preventing HIV Infections and Promoting Safer Sexual Behaviour Amongst Youth in the Rural Eastern Cape, South Africa: Trial Design, Methods and Baseline Findings', *Tropical Medicine & International Health* 11.1: 3–16

Kim, J. *et al.* (2009) 'Assessing the Incremental Effects of Combining Economic and Health Interventions: The IMAGE Study in South Africa', *Bulletin of the World Health Organization* 87.11: 824–32

Kim, J.C. *et al.* (2007) 'Understanding the Impact of a Microfinance-Based Intervention on Women's Empowerment and the Reduction of Intimate Partner Violence in South Africa', *American Journal of Public Health* 97.10: 1794–802

Kirby, D.B.; Laris, B.A. and Rolleri, L.A. (2007) 'Sex and HIV Education Programs: Their Impact on Sexual Behaviors of Young People Throughout the World', *Journal of Adolescent Health* 40.3: 206–17

Leclerc-Madlala, S. (2008) 'Age-Disparate and Intergenerational Sex in Southern Africa: The Dynamics of Hypervulnerability', *AIDS* 22.suppl 4: S17–S25

Luke, N. (2005) 'Confronting the "Sugar Daddy" Stereotype: Age and Economic Asymmetries and Risky Sexual Behavior in Urban Kenya', *International Family Planning Perspectives* 31.1: 6–14

Pronyk, P.M. *et al.* (2006) 'Effect of a Structural Intervention for the Prevention of Intimate-Partner Violence and HIV in Rural South Africa: A Cluster Randomised Trial', *The Lancet* 368.9551: 1973–83

Swann, M. (2018) 'Economic Strengthening for HIV Prevention and Risk Reduction: A Review of the Evidence', *AIDS Care* 30.suppl 3: 37–84

UNICEF (2013) *Towards an AIDS-Free Generation – Children and AIDS: Sixth Stocktaking Report*, New York NY: United Nations Children's Fund

Wingood, G.M. *et al.* (2007) 'Efficacy of an HIV Prevention Program Among Female Adolescents Experiencing Gender-Based Violence', *American Journal of Public Health* 96.6: 1085–90

Wolbers, M. *et al.* (2011) 'Sample Size Requirements for Separating Out the Effects of Combination Treatments: Randomised Controlled Trials of Combination Therapy vs. Standard Treatment Compared to Factorial Designs for Patients with Tuberculous Meningitis', *Trials* 12: 26

This page is intentionally left blank