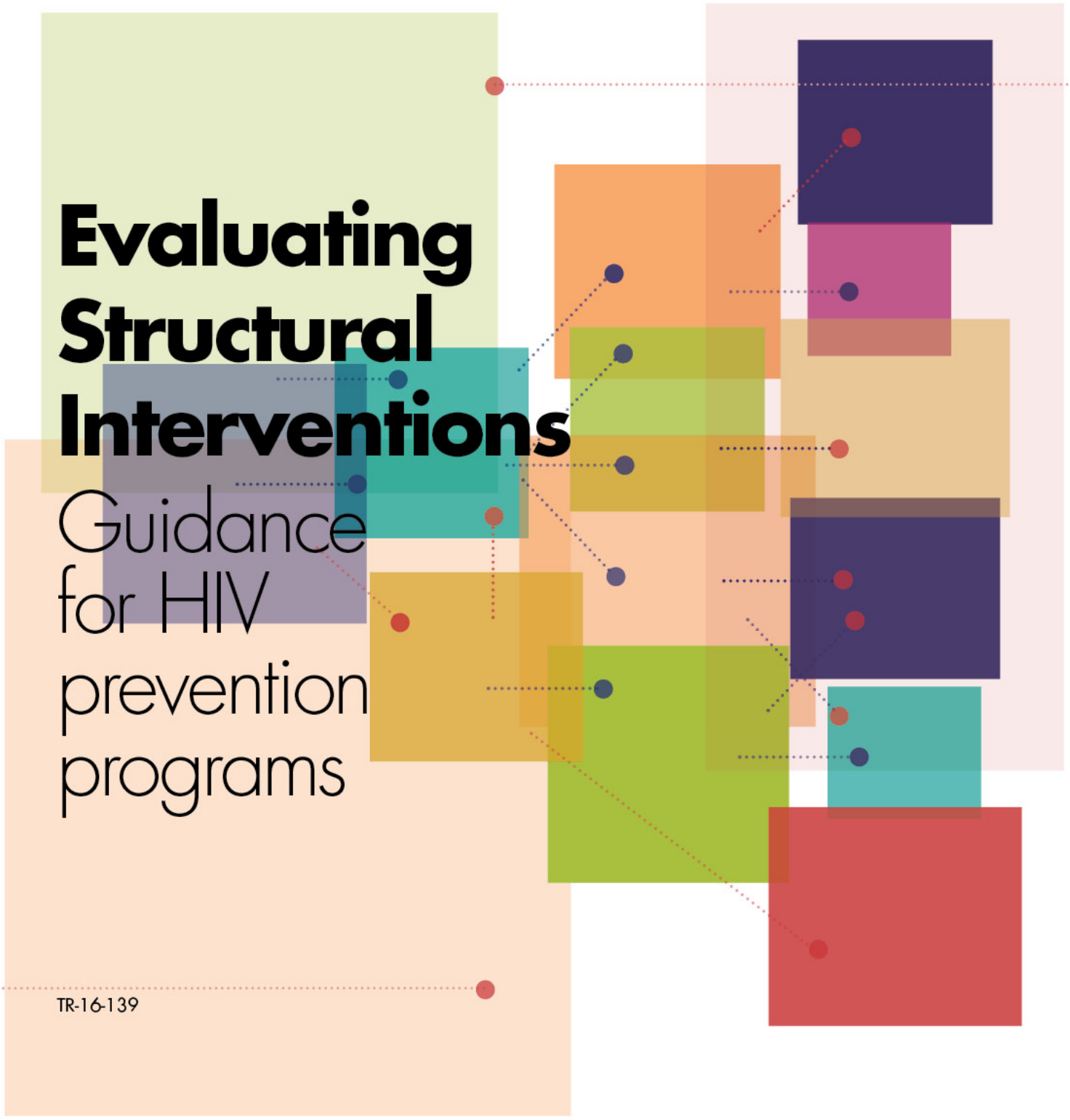
Evaluating Structural Interventions

Guidance for HIV prevention programs



TR-16-139

USAID
FROM THE AMERICAN PEOPLE

PEPFAR

MEASURE Evaluation

Evaluating Structural Interventions

Guidance for HIV prevention programs

ISBN: 978-1-943364-90-9

MEASURE Evaluation
Carolina Population Center
University of North Carolina at Chapel Hill
400 Meadowmont Village Circle, 3rd Floor
Chapel Hill, North Carolina 27517
Phone: +1 919-445-9350 | Fax: +1 919-445-9353
measure@unc.edu
www.measureevaluation.org

This publication was produced with the support of the United States Agency for International Development (USAID) under the terms of MEASURE Evaluation cooperative agreement AID-OAA-L-14-00004. MEASURE Evaluation is implemented by the Carolina Population Center, University of North Carolina at Chapel Hill in partnership with ICF International; John Snow, Inc.; Management Sciences for Health; Palladium; and Tulane University. Views expressed are not necessarily those of USAID or the United States government. TR-16-139

USAID
FROM THE AMERICAN PEOPLE

PEPFAR

MEASURE
Evaluation

ACKNOWLEDGMENTS

This guidance was written by Lauren Hart, Brittany S. Iskarpatyoti, Mahua Mandal, and James C. Thomas.

MEASURE Evaluation wishes to thank those who contributed to the development and publication of these guidelines. First, we would like to acknowledge the United States Agency for International Development (USAID) for its support of this activity, with special thanks to Ana Scholl and Kristen Wares.

We are grateful to those who participated in the key informant interviews and shared their experiences and expertise: Caitlin Kennedy (Johns Hopkins Bloomberg School of Public Health), Clare Barrington (University of North Carolina Gillings School of Global Public Health), Dallas Swendeman (UCLA David Geffen School of Medicine and Fielding School of Public Health), Denise Halfors (Pacific Institute for Research and Evaluation), Hyunson Cho (Pacific Institute for Research and Evaluation), Jhumka Gupta (George Mason University), Kelly Kilburn (University of North Carolina Department of Public Policy), Kim Blankenship (American University Department of Sociology and Center on Health, Risk and Society), Ravi Verma (International Center for Research on Women), Deanna Kerrigan (Johns Hopkins School of Public Health), and Susan Sherman (Johns Hopkins School of Public Health). Their contributions have informed examples, recommendations, and explanations throughout.

Finally, we'd like to thank Jill Lebov for assisting in the literature review; Jessica Fehringer and Carolina Mejia for their technical review; and MEASURE Evaluation's Knowledge Management team for design, editorial, and publication assistance.

CONTENTS

EXECUTIVE SUMMARY

Structural interventions aim to improve health outcomes by altering the social, economic, and legal-political environment in which health processes and outcomes are embedded (Blankenship, et al., 2006; Parkhurst, 2013). By targeting structural factors such as poverty and education, structural interventions aim to influence the underlying context of HIV risk (Blankenship, et al., 2006). Investment in structural interventions and their evaluations has lagged behind that in behavioral and biomedical interventions for HIV prevention. Additionally, methodological challenges in evaluating structural interventions have contributed to the relative dearth of evidence about their effectiveness (Gupta, et al., 2008; Heise & Watts, 2013). In addition to showing that a structural intervention works in a particular context, it is important to provide evidence for why, under what circumstances, for whom, and at what cost the intervention is effective (Heise & Watts, 2013).

The process of evaluating structural interventions is much the same as that of evaluating other interventions. However, in this guidance we highlight strategies and considerations that are uniquely important in the former. We describe the process in six steps:

1. **Engage stakeholders**. Because the implementation and effects of structural interventions span the individual, community, societal, and political levels, representatives from each of these levels should be engaged in order to better understand intended and unintended factors and consequences of a program.
2. **Describe a theory of change**. Structural interventions target distal factors on the assumption that changes in those factors will lead to changes in proximal factors downstream and thus to intended outcomes. A well-defined theory of change helps program designers and implementers think through how that occurs.
3. **Define a research question**. The research question will depend on the outcome of interest, intermediate factors identified in the theory of change, and the type of evaluation that can be supported. Because of the complexity of structural interventions and the need to balance competing interests with limited resources, evaluators should engage stakeholders in developing research questions. These questions may address how various aspects of the program are implemented; whether and how the program contributes to intermediate health outcomes; and/or the program's degree of success in preventing HIV.
4. **Choose a design**. The design for evaluating a structural intervention must adequately answer the evaluation questions while also meeting ethical, feasibility, and other constraints. This can provide a challenge, as can the complexity of the intervention or the breadth of factors that affect or are affected by it.
5. **Analyze and report**. By nature, structural interventions must take into account a web of confounding and intermediate variables, making careful analysis and reporting even more imperative. When an intervention has multiple components, as is often the case in structural interventions, measuring synergies may be required to determine whether two components working together are creating greater change than either could alone.
6. **Replicate and generalize**. Because structural interventions rely heavily on their context, it is challenging to determine which program components are generalizable and which are unique.

Although numerous challenges exist in evaluating structural interventions, current evidence indicates their promise for HIV prevention. By planning for evaluations early, working with stakeholders, determining a theory of change, carefully selecting research questions, and selecting the most appropriate research design, those in the HIV-prevention field can continue to determine which interventions are most effective.

ABBREVIATIONS

FSW	female sex worker
GBV	gender-based violence
IOM	Institute of Medicine
LMIC	low- and middle-income countries
M&E	monitoring and evaluation
MSC	most significant change
PEPFAR	President's Emergency Plan for AIDS Relief
PMTCT	prevention of mother-to-child transmission
PrEP	pre-exposure prophylaxis
RCT	randomized controlled trial
TOC	theory of change
UNAIDS	Joint United Nations Programme on HIV/AIDS
USAID	United States Agency for International Development

PURPOSE OF THIS GUIDANCE

This guidance is intended to assist program planners, managers, and implementers in understanding the basics of and planning for evaluations of structural interventions for HIV prevention. It is intended to guide staff members at donor and implementing organizations working with researchers and evaluators to understand the steps required in planning an evaluation, and it is a starting point for identifying research questions, considering design options, reporting results, and addressing the question "So what?" This guidance should be used along with other evaluation resources, such as those included throughout this document. It is not intended to serve as a complete resource on evaluation design, data collection, analysis, and use.

DEVELOPMENT OF THIS GUIDANCE

This guidance was developed by MEASURE Evaluation and informed by (1) a systematic review of peer-reviewed journal articles, (2) a desk review of grey literature, and (3) interviews with researchers and evaluators who have expertise in conducting evaluations of structural interventions in low- and middle-income countries (LMICs).

We conducted a systematic review to examine the methods used to evaluate the outcome and impact of structural interventions aimed at HIV prevention through economic empowerment, formal and informal education, and reduction of substance use in LMICs. (See Appendix A for details.) The desk review had a similar focus, but it included grey literature beyond outcome and impact evaluations. We conducted a targeted search of websites for organizations known to be involved in evaluating structural interventions for HIV prevention. Documents reviewed included reports, evaluation plans, and working papers.

To understand how evaluations were carried out and what lessons have been learned, we conducted key informant interviews with eleven experts from the field (see Acknowledgements). Initial contact was made with authors identified in the systematic review, and subsequent snowball sampling was used to identify additional experts.

INTENDED USERS

This guidance is written for donors and national governments interested in evaluating structural interventions; HIV program planners, managers, and implementers of structural interventions; M&E managers and officers tasked with monitoring structural interventions; and institutes or agencies tasked with external evaluations of structural interventions. It is also applicable to all LMICs developing or implementing structural interventions to prevent HIV.

INTRODUCTION

What Are Structural Interventions?

Structural interventions aim to improve health outcomes by altering the social, economic, and legal-political environment in which health processes and outcomes are embedded (Blankenship, et al., 2006; Parkhurst, 2013). In HIV prevention, individuals' attitudes, decisions, and behaviors are influenced by a range of contextual factors, including economic and social conditions, that may contribute to their vulnerability to HIV infection (Gupta, et al., 2008; Parkhurst, 2013; Pronyk & Lutz, 2013). By targeting structural factors such as poverty and education, structural interventions aim to influence the underlying context of HIV risk (Blankenship, et al., 2006; Tumlinson, et al., 2015). Some development professionals and academics have suggested that efforts at HIV prevention have not kept pace with HIV care and treatment, owing to insufficient attention to HIV's structural factors.

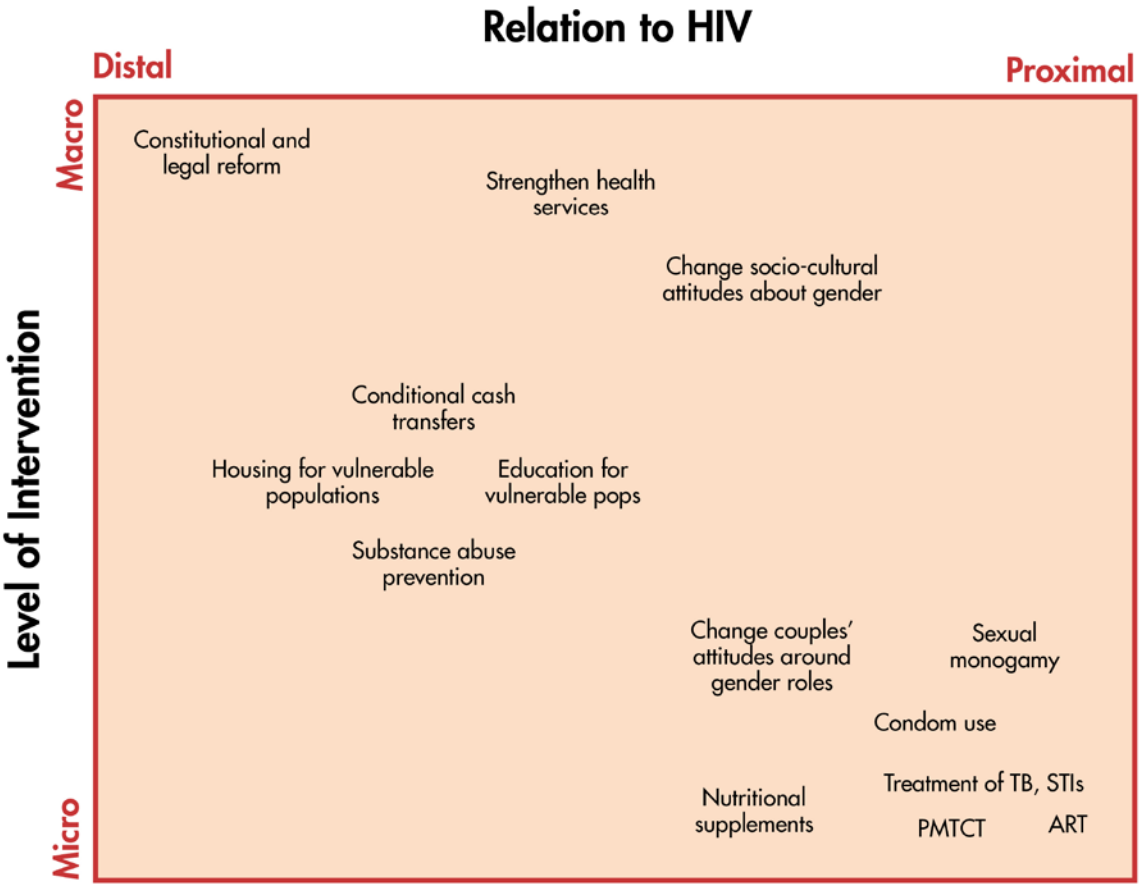
Interventions that target structural factors can have a wide impact and lead to sustainable reductions in HIV risk and vulnerability (Heise & Watts, 2013). Intervening in just one structural factor can decrease the risk of infection in a number of ways (Heise & Watts, 2013). For example, poverty is a commonly cited social "structure" that may increase women's risk of gender-based violence (GBV), leading to their inability to negotiate condom use and an increase in their risk of contracting HIV (Krishnan, et al., 2008; Heise & Watts, 2013). Poverty increases the risk of heavy alcohol use, which may influence partner selection and risky sexual behaviors (Heise & Watts, 2013). Interventions that work to increase economic empowerment can therefore affect vulnerability to HIV in multiple ways.

Structural interventions may be applied in combination with behavior change and/or biomedical interventions—often called combination interventions (Gupta, et al., 2008). The President's Emergency Plan for AIDS Relief (PEPFAR) 3.0 promotes combination interventions, noting the importance of "bringing together many relevant approaches from multiple sectors—education, health, community, economic and psychosocial—to establish a core package of evidence based interventions" (The Office of the U.S. Global AIDS Coordinator, 2014). For example, an HIV-prevention combination intervention aimed at female sex workers and their clients may include a behavioral component that targets individuals' risky sex behavior through education around condom use; a biomedical component that provides pre-exposure prophylaxis (PrEP) to female sex workers; and a structural component that provides livelihood training to female sex workers as an alternative to sex work (Pronyk & Lutz, 2013).

Levels of HIV Intervention

HIV intervention can be designed at various levels of implementation, from the individual (micro) to the societal (macro) and everything in between. Evaluations that informed this guidance are largely based on interventions that act on HIV risk factors by influencing economic opportunity, formal and informal education, and substance use and abuse—that is, interventions that are considered more macro than micro. These interventions also target factors that are distal or remote from HIV infection. (See Figure 1.)

Figure 1: Common HIV interventions by level of intervention and their relation to HIV outcomes



Role of Structural Interventions within Development Priorities

The importance of structural factors in determining level of risk, and the need for interventions targeting these factors, are becoming more broadly recognized by researchers, funders, and policymakers (Auerbach, Parkhurst & Cáceres, 2011; Gupta, et al., 2008; Pronyk & Lutz, 2013; The Office of the U.S. Global AIDS Coordinator, 2014). This view of HIV prevention is part of a global shift away from disease-specific priority setting and could favor interventions that influence a wider range of structural factors (Pronyk & Lutz, 2013). As an example, structural approaches are embedded within two of the three pillars of the Joint United Nations Programme on HIV/AIDS (UNAIDS) investment framework: critical enablers and synergies with development sectors (Pronyk & Lutz, 2013). PEPFAR 3.0 also stresses the importance of structural factors: "PEPFAR's dedication to addressing HIV/AIDS through access to prevention, care, and treatment means removing barriers to health services for all people, but in particular those most vulnerable who often disproportionately suffer from cultural and structural obstacles to care" (The Office of the U.S. Global AIDS Coordinator, 2014).

The importance of addressing structural factors is also included in recommendation 5-1 of the 2013 Institute of Medicine (IOM) evaluation of the PEPFAR program, which states, "To contribute to the sustainable management of the HIV epidemic in partner countries, PEPFAR should support a stronger emphasis on prevention. The response should incorporate an approach balanced among biomedical, behavioral, and structural interventions that is informed by epidemiological data and intervention

effectiveness evidence. PEPFAR should support advances in prevention science to expand the availability of effective interventions where knowledge is lacking" (Institute of Medicine, 2013).

Despite the promise of structural interventions and donor enthusiasm for additional efforts in their implementation and evaluation, less data has been collected on structural interventions than on biomedical and behavioral interventions, and few existing programs have been rigorously evaluated against biological outcomes, such as HIV biomarkers (Gupta, et al., 2008; Heise & Watts, 2013; Pronyk & Lutz, 2013).

Importance of Evaluating Structural Interventions

Investment in structural interventions and their evaluations have lagged behind that in behavioral and biomedical interventions for HIV prevention. Additionally, methodological challenges in evaluating structural interventions have contributed to the relative dearth of evidence about their effectiveness (Gupta, et al., 2008; Heise & Watts, 2013). Given these factors, it is vital that structural interventions for HIV prevention be evaluated (Sommer & Parker, 2013).

Donors and researchers alike have stressed the need for additional research. The IOM evaluation of PEPFAR 2.0 noted the disproportionately low availability of monitoring data and evidence from rigorous evaluations for structural interventions in comparison with that for biomedical interventions, and identified a "critical need for improved application of advances in social and behavioral science–based research and evaluation science for prevention to determine the most effective combination of prevention interventions in diverse country contexts" (Institute of Medicine, 2013).

In addition to showing that a structural intervention works in a particular context, it is important to provide evidence for why, under what circumstances, for whom, and at what cost the intervention is effective (Heise & Watts, 2013).

Setting Realistic Expectations for Evaluations

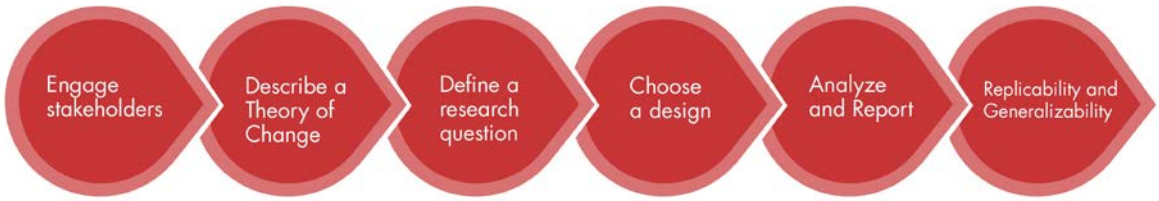
Structural interventions are difficult to evaluate for a number of reasons. (1) Their contextual nature means that factors such as economic barriers, political and legal systems, cultural norms, and power dynamics influence how an intervention is implemented by program staffers and received and taken up by local communities (Campbell, et al., 2000). This influences the variables that must be taken into account when designing and implementing an evaluation. (2) The wide range of contextual factors across settings has implications for the external validity of an evaluation—or the degree to which evaluation results from one context can be generalized to another context. (3) Structural interventions often comprise many parts that interact with one another as well as with the built and social environment (Craig, et al., 2008). Evaluators must make assumptions about how these factors are interrelated and determine directions of "causality." (4) Implementation of structural interventions may be non-linear, iterative, and adaptive (Campbell, et al., 2000), making it challenging to use conventional evaluation methods such as testing discrete hypotheses and measuring a predetermined and limited number of intermediate variables and outcome variables. (5) Random assignment of groups may be infeasible or unethical (Bonell, et al., 2006), making it difficult to rule out selection bias or to select an appropriate control group, and thus to conclude whether the intervention led to changes in health outcomes or the exposed group was already more prone to improving its health. (6) Structural interventions aim to influence factors that are "upstream" from health outcomes. As a result, measurable changes in health outcomes and health status may not be detectable within the relatively short timelines of government and donor project cycles (Sommer & Parker, 2013).

Our intention in enumerating these challenges is to assist donors, governments, and partners in understanding the importance of setting realistic expectations for structural-intervention evaluations rather than to suggest that such evaluations are not useful. For example, donors may be interested in

funding a randomized controlled trial (RCT), which is considered the gold standard in biomedical research and, more broadly, in public health. However, RCTs may be infeasible or inappropriate for evaluating certain types of structural interventions (Bonell, et al., 2006; Thomas, Curtis, & Smith, 2011). Evaluations must be simultaneously rigorous and flexible, and evaluators should note and account for the varied contexts, indirect pathways, and potential synergies that might influence evaluation results.

GUIDANCE FOR EVALUATING STRUCTURAL INTERVENTIONS FOR HIV PREVENTION

This guidance provides recommendations and strategies for evaluating structural interventions to prevent HIV transmission. Although the process is much the same as that for evaluating other interventions, we will highlight strategies and considerations that are uniquely important when evaluating a structural intervention. There are six steps in the process:



These steps do not always follow a linear progression like the one above, and they are often completed through an iterative process. Evaluators frequently need to revisit a step or to complete other steps concurrently. Regardless of where one starts, each step informs the next step or series of steps.

Engage Stakeholders

As with other program evaluations, one of the first steps in evaluating a structural intervention is to identify and engage stakeholders. The appropriate set of stakeholders can shed light on the multitude of factors that should be considered in the evaluation, help identify appropriate control groups, and maximize the impact of data ownership and use for real-world benefit. Stakeholders may include program managers and staffers, national and regional government representatives, partner organizations, clients/service recipients, funders, community representatives, and volunteers.

Because the implementation and effects of structural interventions span the individual, community, societal, and political levels, it is necessary to engage representatives from each. A concerted effort should be made to engage a wide set of stakeholders to better understand the intended and unintended factors and consequences of a program. We recommend working with partners that are already established on the ground and are familiar with and can identify other government agencies, organizations, or community groups. Stakeholder engagement tools can be used to determine and prioritize groups that implement the program, are served by the program, and/or would be interested in the results of the program evaluation (MEASURE Evaluation, 2011).

Once partners have been identified, ongoing dialogue with them is important—particularly discussions that highlight the varying priorities of each stakeholder and ways to narrow a wide range of interests. Evaluators also need to gain the trust of program implementers and of the population the intervention is trying to address. If implementers fear they will be criticized for how they are running their program, or members of the target population are afraid to reveal unhealthy behaviors, they will be less apt to share valid information with evaluators.

For additional information on engaging stakeholders, see the MEASURE Evaluation Stakeholder Engagement Tool.

Describe a Theory of Change

Even if not required by a funding source, an important step in evaluating structural interventions is to develop a theory of change (TOC)—defining the causal pathways through which an intervention is expected to achieve its impact within the particular context (DeSilva, et al, 2014). A TOC provides a shared understanding of how structural, behavioral, and biomedical factors and programming pieces are expected to play a role in reaching the desired results, and where each factor lies on the pathway to change.

Unlike behavioral or biomedical interventions, which are likely to target proximal factors, structural interventions target more-distal factors. The assumption is that changes in them will lead to changes in proximal factors downstream via intermediate factors in the causal pathway. Well-defined theories of change help program designers and implementers think through how this will occur.

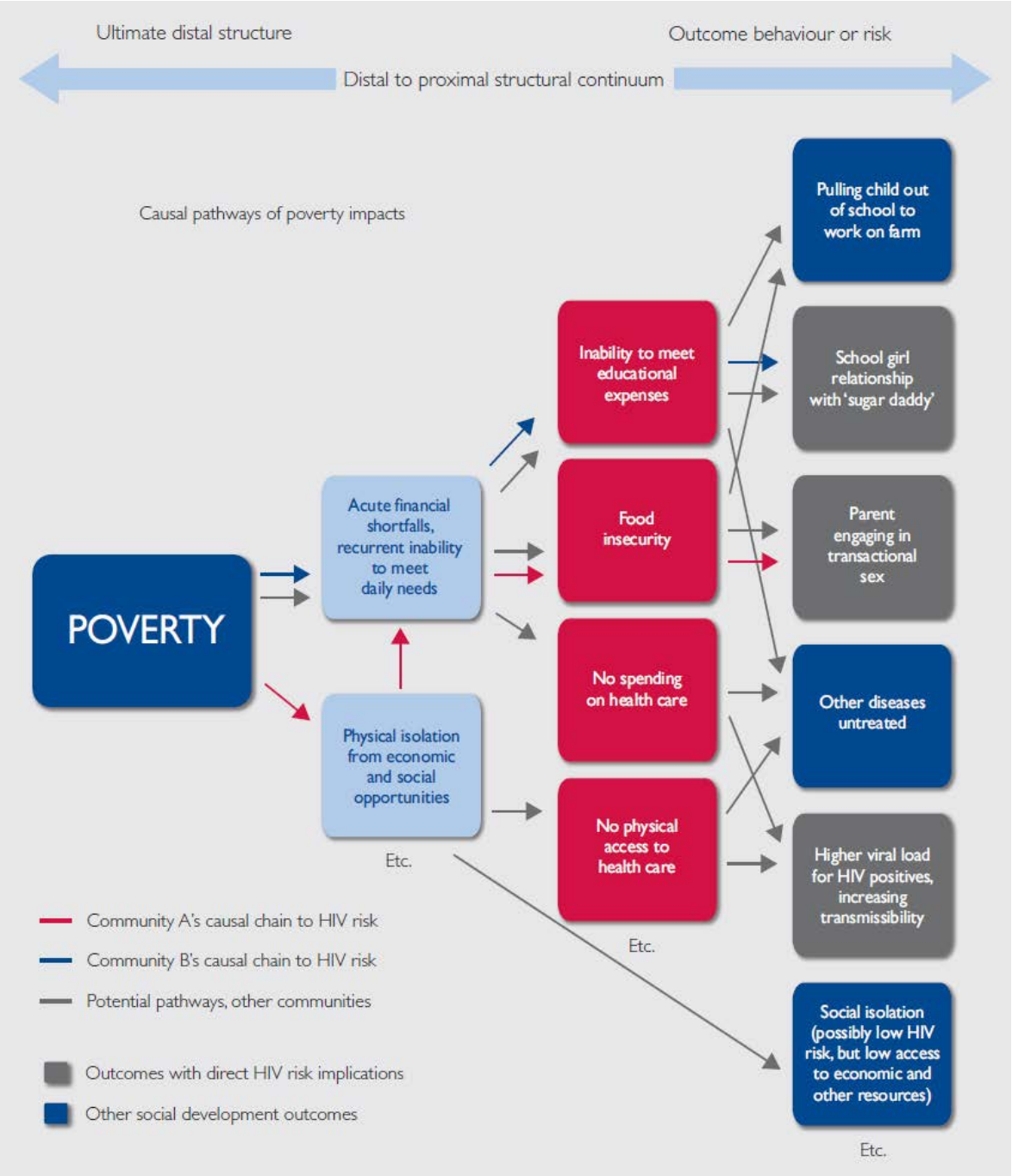
For example, an intervention that provides conditional cash transfers to students is intended to increase access to education (a distal factor) in order to ultimately decrease HIV incidence (the outcome of interest). The theory of change should outline the intermediate factors that are believed to exist between students' receiving cash transfers and their risk of HIV; each link between steps should be grounded in evidence or theory.

A TOC may include only those factors that are explicitly part of the intervention, as in the example above. However, TOCs of structural and other complex interventions often include factors that lie outside the specific intervention (see Figure 2). Given the nature of structural interventions, we advise including factors in the social environment that are theorized to influence program outcomes, even if they are not among the factors the intervention seeks to alter. All the factors included in the theory of change are important in informing indicator selection. While the overall goal of the evaluation question is to understand whether or not an intervention affects the outcome of interest, an evaluation may also investigate questions about other factors included. This is especially important when evaluating structural interventions, because they are likely to work through several, potentially complex pathways (Bonell, et al., 2006).

Determining Timelines

Researchers and evaluators of structural interventions have raised the importance of timing in data collection in two ways. First, they suggest allowing more time between the start of the program and the collection of endline data. Second, if feasible, they recommend collecting data at multiple points throughout the program. Collecting data in such ways would allow interventions that act on structural factors sufficient time to produce impact-level results.

Figure 2. Example of causal pathways from poverty to HIV risk



Source: Parkhurst, Justin O. (2013). Structural drivers, interventions, and approaches for prevention of sexually transmitted HIV in general population: Definitions and an operational approach. Structural Approaches to HIV Prevention Position Paper Series. Arlington, VA: USAID's AIDS Support and Technical Assistance Resources, AIDSTAR-ONE, Task Order I, and London: UKaid's STRIVE Research Consortium.

Additionally, when it is impossible to include a sample size sufficiently large to test for a meaningful and statistically significant change in the primary outcome, the study may examine changes in the intermediate factors. For example, a study intended to measure the impact of a structural intervention on reducing women's experience of gender-based violence and new cases of HIV infection might find that powering for those outcomes would require prohibitively large sample sizes. However, a review of the evaluation's TOC could reveal that changing perceptions of rigid gender roles leads to both decreased acceptance of

GBV and increased ability of women to negotiate condom use. These outcomes, in turn, lead to less GBV and fewer cases of new HIV infection. Under this TOC, the evaluation might instead power the study on the percentage of women and men who accept GBV as a way to discipline women and the percentage of women who report that they are able to negotiate condom use for every act of sex.

If no theory of change exists at the start of an intervention, the evaluation team must gather information from program designers (in the form of written documents such as proposals and work plans, and through discussions) and consult other stakeholders to outline a TOC. The adaptive nature of well-implemented structural interventions also speaks to the flexibility evaluators should have to refine predetermined TOCs on the basis of monitoring and midline data.

For more information on developing a theory of change, visit the resources hyperlinked below:

Review of the use of 'Theory of Change' in international development.

http://www.theoryofchange.org/

Define a Research Question

Once evaluators have described a theory of change for the intervention and understand how the intervention can affect health outcomes, an appropriate research question will be needed. Evaluators should engage stakeholders to identify the questions they want answered regarding how various aspects of the program are implemented (process evaluation); whether and how the program contributes to intermediate health outcomes such as HIV knowledge, attitudes, and behavior (outcome evaluation); and/or the program's degree of success in preventing HIV (impact evaluation).

The research question will depend on the outcome of interest, intermediate factors identified in the TOC (see sidebar), and the type of evaluation that can be supported. Process evaluations are less expensive than outcome and impact evaluations, and may be implemented relatively quickly; however, they typically assess how the program was implemented (e.g., the types and quantities of services delivered and the beneficiaries of those services) and whether it was implemented as planned. Process evaluations may also include examining the resources used to deliver services, the challenges encountered, and the ways in which those challenges were resolved. However, to address questions about the program's effect on HIV requires an outcome or impact evaluation, for which greater capacity and resources are necessary.

Strategies for choosing a research question include meeting with stakeholders and interviewing key informants, such as members of the intervention implementation team. These meetings should be part of the stakeholder dialogue discussed above. Because of the complexity of structural interventions, the number and diversity of stakeholders, and the need to balance competing interests with limited resources, this may take considerable time and effort. However, ensuring that all stakeholders are aware of the scope and limitations of a study is essential to securing satisfaction with and use of the results.

Take, for example, an evaluation of an intervention to prevent HIV that seeks to measure its impact on education and HIV outcomes. Key components of the intervention include conditional cash transfers for adolescent girls who graduate from senior secondary school and small-group life skills and HIV prevention education sessions. The intervention also includes numerous parental and community engagement activities, to help ensure that the program is accepted and supported by the community. Research questions pertinent to this impact evaluation might include:

- Does the full package of services (conditional cash transfer + small-group education) improve education and health outcomes by decreasing school absence, increasing school retention,

increasing graduation rates, increasing use of voluntary counseling and testing services, and decreasing risky sexual behaviors?

- Do adolescent girls who receive only the conditional cash transfer improve their school attendance and progression, reduce their risky sexual behaviors, increase their use of voluntary counseling and testing services, and decrease their probability of contracting HIV?
- Do adolescent girls who receive only the small-group education reduce their risk of sexual behaviors, increase their use of voluntary counseling and testing services, and decrease their probability of contracting HIV?

To better understand the enabling environment, the evaluation might also include these questions:

- To what degree are adolescents' perspectives on the parental support they received congruent with parents' perspectives on the support they provided?
- How do various aspects of parental and community support influence adolescents' educational and health processes and outcomes?

Note that, as mentioned under Describing a Theory of Change, research questions may be developed for multiple factors in the TOC. It may also be necessary to include questions about process along with questions about outcome or impact, because implementation is likely to vary according to context, and evidence about the process may illuminate the relationship between context, delivery, and outcomes. This information can help in determining the generalizability of the intervention (Bonell, et al., 2006). We will discuss this idea of generalizability further in the section: Replicate and Generalize: Addressing the Question "So What?"

Balancing Research Priorities with Intersectionality

As we've discussed, structural interventions often target distal factors such as lack of access to education or economic opportunity—which are shaped by the intersection of social, cultural, racial, religious, and gender inequities. How these factors are included in the evaluation approach, data analysis, and reporting is significant, because "inequities are never the result of single, distinct factors. Rather, they are the outcome of intersections of different social locations, power relations and experiences" (Hankivsky, 2014). A comprehensive TOC can illustrate the interconnectedness of these elements and how they compound into systems of discrimination or disadvantage within a given population.

Once included in a TOC, these factors can be measured intentionally. Mixed methods and other designs noted below can help to clarify and prioritize elements within evaluations that incorporate an intersectionality perspective.

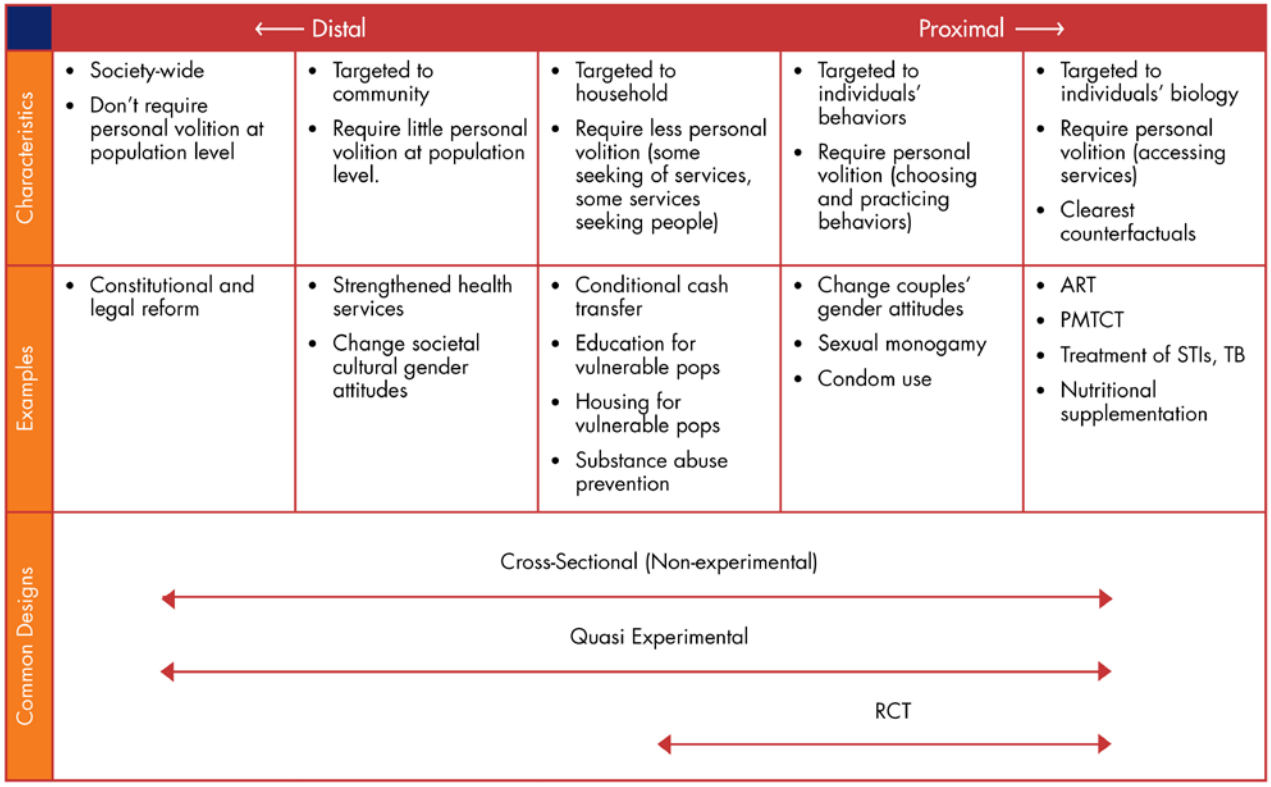
Choose a Design

When evaluating a structural intervention, choosing a design that adequately answers the evaluation questions and also meets other constraints can be challenging. The challenge may lie in the complexity of the intervention or the breadth of factors affected by it. And as with any evaluation, budget, time, data, and political constraints will also apply (Bamberger, Rugh & Mabry, 2011). It is also important to identify the target audience for the evaluation results. Effort should then be made to engage these stakeholders and learn what results will be acceptable for decision making (Bonell, et al., 2006; Habicht, Victora & Vaughan, 1999).

Determining the research design may be an iterative process. It should involve gathering information on the nature of the intervention, the affected population, and other important background information, and should include stakeholders.

A number of design options allow evaluators to answer whether and how a structural intervention "works" while simultaneously addressing the numerous constraints that arise in evaluating complex interventions. In the following sections, we will look at methods and designs and why they are or are not appropriate for evaluating structural interventions.

Table 1. Review of HIV interventions and common evaluation designs



Approaches for Evaluating Structural Interventions

Mixed Methods

When we discussed evaluating structural interventions with experts in the field, mixed methods came up again and again as essential to understanding the complex factors that influence structural interventions. Mixed-methods evaluations are those that use more than one method of data collection—often quantitative and qualitative methods either simultaneously or sequentially. Reasons for choosing a mixed-methods design include (1) to inform the development of data collection instruments (e.g., using focus groups to help design a survey); (2) to collect complementary information that allows the researcher to compare findings from multiple sources (see the box on triangulating data); (3) to collect information from people or groups whose input might otherwise be excluded; (4) to explain quantitative results—providing examples of and insight into unanticipated changes or responses or otherwise aiding in interpretation; and (5) to allow for flexibility in design (Hall & Roussel, 2012; USAID, 2010). Different information is derived from different methods. Collecting qualitative information from a smaller sample of the quantitative cohort at multiple points can add rich information across the life of a project.

The mix of methods included in an evaluation will be determined by the evaluation questions. Using mixed methods can help (1) guide the rest of the evaluation design; (2) understand what it is like to be a participant in a particular evaluation; (3) explain the "why" or "how" of changes identified from quantitative data collection; (4) gather information from hard-to-reach populations; (5) make changes to study design; and (6) identify otherwise-unidentified errors in the quantitative data collection.

For more information on using mixed methods as part of an evaluation, see the resources hyperlinked below:

- Performance Monitoring and Evaluation TIPS: Conducting Mixed-Method Evaluation by USAID
- Designing and Conducting Mixed Methods Research

Triangulating Data

Triangulation is the process of gathering and comparing data from more than two sources. These may include questionnaires, focus groups, vital registration data, demographic health survey data, household surveys, or key informant interviews. If the data from all sources supports the same outcome or impact, the researchers' conclusions have more certainty. If different sources provide different or potentially conflicting information about program outcomes or impact, the researchers must investigate further. They may end up questioning the validity of the data from one or more sources, or they may realize that the outcome or the impact is more complex than it appeared from only one data source (BetterEvaluation, 2014; USAID Asia, 2010).

Especially when dealing with complex problems, triangulation can be an important tool. Whether or not the data sources confirm one another, it leads to a deeper understanding of the situation and ultimately a greater confidence in the researchers' conclusions.

For more information on triangulating data, visit the following resources:

- Triangulation by BetterEvaluation
- Triangulation by the Robert Wood Johnson Foundation
- An Introduction to Triangulation by UNAIDS
- Data Triangulation for HIV Prevention Program Evaluation in Low and Concentrated Epidemics by USAID Asia

Complexity-Aware Monitoring

Complexity-aware monitoring has been identified as helping to better understand complex interventions. The United States Agency for International Development suggests using this process with strategies or projects in which cause-and-effect relationships are not well understood (USAID, 2013). Structural interventions that rely on adaptive management and/or seek to influence social change are good candidates for complexity-aware monitoring. The approaches used in complexity-aware monitoring can also be applied to the design of evaluations for complex interventions.

USAID recommends five approaches to complexity-aware monitoring: (1) sentinel indicators, (2) stakeholder feedback, (3) process monitoring of impacts, (4) most significant change, and (5) outcome harvesting. Sentinel indicators and process monitoring of impacts both work to capture information about predicted results. (For more information on those techniques, see the USAID Discussion Note: Complexity-Aware Monitoring.) Stakeholder feedback is an important type of qualitative information. (More information about including qualitative methods in the evaluation of structural interventions is included in the discussion of specific designs below.) Most significant change and outcome harvesting are both discussed in more detail below.

For more information on implementing complexity-aware monitoring as part of an evaluation, visit:

- Discussion Note: Complexity-Aware Monitoring Discussion Note by USAID

Developmental Evaluation

Developmental evaluation is designed for use in complex or uncertain environments (BetterEvaluation, n.d.-a). The evaluation team shares information with the implementers throughout the intervention, and this feedback is used to adapt implementation. Information flows between the two teams throughout the process.

Because structural factors often operate in complex ways, developmental evaluation may be appropriate for evaluating a structural intervention. It can also be useful when the implementation team is wary of being judged or "graded" for its performance, because the evaluation and implementation teams work hand in hand. However, developmental evaluation is most appropriate when the main motivation is to innovate and adapt the intervention rather than a rigorous external publication of results (BetterEvaluation, n.d.-a).

For more information on implementing developmental evaluation, visit:

- Developmental Evaluation by BetterEvaluation

Selected Methods for Evaluating Structural Interventions

Randomized Controlled Trials

An RCT is a research study in which participants are assigned to treatment/intervention conditions at random (i.e., they have an equal probability of being assigned to the experimental or control group) (BetterEvaluation, n.d.-b). Randomization at the level of a community or larger group is called cluster randomization. The randomization in these studies, which compare an intervention group to a control group (e.g., no intervention, usual care, or another intervention), typically at two or more points in time, allows for the impact to be attributed to the intervention.

Although RCTs are considered the gold standard in biomedical and traditional public health research, implementing them in the context of a structural intervention may pose several challenges, including potential ethical, cost, and feasibility issues, such as difficulty in identifying a suitable control group or counterfactual; potential for contamination of the control group through inadvertent exposure to the intervention; a high cost of implementation; and a high level of capacity needed to implement the evaluation.

These challenges are not unique to structural interventions, but their severity is often magnified when the unit of analysis is a group rather than an individual. For example, in the evaluation of a rural community mobilization project in Tanzania, the randomization was successful because each community was highly isolated. However, when evaluating a similar intervention in more urban settings, the same researcher found that these communities were more porous, allowing members and information to pass easily between them and leading to contamination of the control sites. That is, individuals in the control sites received the information and other services that were being evaluated. In cases where it would be unethical to withhold the intervention from the control group, evaluators should seek another method or determine a way to offer the intervention to the control group in the future. And when there are cost constraints, the evaluator can choose either to pursue another evaluation design or to advocate for additional funds for an RCT.

Many feasibility issues can be addressed by engaging with stakeholders. For example, for an evaluation of a government program that will be implemented at the district level, the government may be willing to take a phased approach to rolling out the program (termed a "step-wedge" design), thus allowing for randomization by district. Stakeholders can shed light on the level and type of information transfer that may occur between communities.

Despite the challenges in using an RCT to evaluate a structural intervention, some research questions can be answered only with this method. And governments and donors, especially in academic circles, are more likely to consider findings from an RCT favorably (Cook, 2002).

Finally, even when an RCT is possible, including qualitative methods in the evaluation design can add richness to the results (Kennedy, 2015).

For more information on RCTs, visit the resources hyperlinked below:

- Randomized Controlled Trial (RCT) by BetterEvaluation
- An introduction to the use of randomized control trials to evaluate development interventions; 3ie working paper
- Randomized Controlled Trials (RCT) by UNICEF

Quasi-experiments

Like RCTs, quasi-experimental methods test a causal hypothesis, meaning that they investigate whether an intervention leads to expected results. Unlike RCTs, however, quasi-experimental methods use a comparison group, which differs from a control group in that it is not randomly selected. In quasi-experimental evaluation designs, whether or not a community receives the intervention is often the result of programmatic or policy decisions made by governments, donors, and program designers and implementers. Without the benefit of randomization, such evaluations must attempt to control or otherwise account for factors unrelated to the intervention or characteristics of the communities receiving the intervention. Such factors may strongly influence the outcomes of interest, thus making it difficult or impossible to attribute changes to the intervention.

Quasi-experimental methods are often used to evaluate structural interventions because of ethical and practical considerations in randomizing groups. For example, evaluating structural interventions that target positive social factors such as economic empowerment and education for HIV prevention is particularly challenging: Receipt of the intervention is inarguably beneficial, but whether it prevents HIV infection is unknown. Therefore, randomizing recipients impacts more than the evaluative outcomes. More on ethical considerations can be found at the end of this guide.

Quasi-experimental designs come in many types, with a variety of applications in specific contexts. For more information on implementing quasi-experimental methods as part of an evaluation, visit:

- Quasi-Experimental Design and Methods by UNICEF

Regression Discontinuity

Regression discontinuity is an evaluation strategy that can be used to reduce the effects of selection bias and unequal intervention and comparison groups in quasi-experimental designs. This is particularly important for programs or policies that use a specific criterion that is associated with the outcome of interest to determine whether individuals or communities are eligible to receive the intervention. For example, an intervention that provides cash transfers to adolescent girls who complete senior secondary school may limit eligibility to those from the poorest families in the district—but children from the poorest households are the least likely to graduate from senior secondary school. If members of the comparison group were identified by randomly selecting adolescents in the district not receiving the intervention, the comparison group would most likely have a higher average household income (compared with the intervention group) and thus would be more likely to graduate. Equal study groups—that is, groups whose members have approximately the same household income—could be identified by using a narrow range around the intervention criterion ("cutoff") for household income in both the intervention group and the comparison group. The assumption is that adolescents from households just above the income cutoff (who would be in the comparison group) or just below the income cutoff (who would be in the intervention group) are similar; and thus an evaluation comparing the two groups would **provide the treatment effect.**

For more information on implementing quasi-experimental methods as part of an evaluation, visit the following resources:

- Module 12: Advanced Evaluation Designs (IEs). USAID
- Regression Discontinuity by Better Evaluation
- A Practical Guide to Regression Discontinuity by MDRC

Cross-Sectional

A cross-sectional study assesses the characteristics of a population at one point in time in which data is collected from a sample (ideally representative) of the population. A structural-intervention evaluation might use a cross-sectional survey to examine the association between health outcomes of interest and exposure to the program in an intervention area only (Alexander, et al., n.d.). Such evaluations might use two or more cross-sectional surveys, implemented at baseline (before the start of an intervention), at endline (at the end), and possibly at midline (in the middle), allowing for an examination of changes in key indicators over time. Cross-sectional surveys before and after policy implementation permits examination of the effects of national policies on health outcomes. Finally, these surveys may be implemented at multiple points in an RCT or a quasi-experimental evaluation design in both intervention and control or comparison groups. In such cases, a representative sample of the population would be surveyed at each point, allowing researchers to estimate the average change over time within the populations.

A cross-sectional study design may be chosen for its low cost and—if it is implemented only once—relatively quick time frame. Additionally, using cross-sectional surveys in intervention sites removes the ethical dilemma inherent in an RCT: There is no treatment and no control group, so no one is denied the intervention (Mann, 2003). The main disadvantage to a cross-sectional study is that it is difficult to determine whether the intervention has led to the desired outcome or there is simply an association between the intervention and outcomes (Mann, 2003).

For more information on cross-sectional studies, visit:

- Cross-Sectional Studies; UNC School of Public Health

Most Significant Change

The most significant change (MSC) is a qualitative participatory method that "involves assessing the changes and impacts that have happened as a result of a program from the perspective of program participants" (Lennie, 2011). In its purest form, stakeholders—including participants—are engaged throughout the entire process, involved in determining what should be documented and in analyzing the data. Stories of impact and outcomes are generated at the field level, and a systematic process is used to determine which of them reflect the most significant change (Davies & Dart, 2005). MSC can be particularly useful when evaluating programs that adapt to different or evolving contexts, thus leading to differences in implementation and outcomes. This is often the case with structural interventions.

While MSC is an important technique to consider as part of an evaluation design, it is often most informative as a supplement to other methods. That's because of inherent biases in the MSC process, in which the changes that are considered most significant may be the ones that were easiest to attain or most favorable to the storyteller, rather than those most closely related to HIV prevention or level of risk.

The MSC technique has other benefits for structural interventions, including fostering a sense of ownership of the evaluation process among stakeholders.

For more information on implementing the MSC method in an evaluation, visit the resources hyperlinked below:

- The 'Most Significant Change' Technique - A Guide to Its Use by BetterEvaluation
- The Most Significant Change technique: A manual for M&E staff and others at Equal Access

Outcome Harvesting

Outcome harvesting is a qualitative process in which evidence of outcomes is collected and then analyzed to determine whether or not the intervention in question contributed to those outcomes (Wilson-Grau, 2015; Wilson-Grau & Britt, 2013). Outcomes, which are defined behaviors in this method, may include those exhibited by individuals or groups, institutions or organizations, or whole communities (Wilson-Grau & Britt, 2013). Because outcome harvesting focuses on outcomes rather than on activities or outputs (Wilson-Grau & Britt, 2013), this method should be used when the primary motivation for an evaluation is to examine behavior changes.

Outcome harvesting can be useful in situations where the structural intervention's theory of change has not been verified (Wilson-Grau & Britt, 2013).

For more information on implementing outcome harvesting, visit the resources hyperlinked below:

- Outcome Harvesting by BetterEvaluation

- Outcome Harvesting by the Ford Foundation

Social Network Analysis

Social network analysis can be used to evaluate structural interventions that influence interconnections between individuals or organizations. According to the Robert Wood Johnson Foundation, "social network analysis can help evaluators to:

- Understand the network embedded within a program or initiative, in terms of its density, connectedness, balance, and/or centralization
- Identify subsets within the network, such as cliques or key nodes
- Identify important characteristics about the actors in the network, such as gatekeepers and isolates (outliers with few connections)
- Measure the degrees of centrality and similarity of actors within the network" (Fredericks & Carman, 2013)

This information is especially useful when the evaluation examines questions regarding which individuals or units are involved in the intervention—stakeholders, clients, health facilities, etc.—and their level of interconnectedness. For example, evaluations of structural interventions often include work with hard-to-reach groups (such as people who inject drugs) or closed or semi-closed groups (such as youths attending school), when understanding the network is an important part of understanding how the intervention, which may include the diffusion of ideas, innovations, or actions, is functioning. At the cluster level organizational network analysis can be used to examine the interconnectedness between facilities and organizations providing referrals for HIV services (Thomas, et al., 2016).

For more information on implementing social network analysis, visit the resources hyperlinked below:

- Using Social Network Analysis in Evaluation by the Robert Wood Johnson Foundation
- Using Social Network Analysis for M&E by BetterEvaluation
- MEASURE Evaluation Summary of Approaches: http://www.cpc.unc.edu/measure/resources/publications/sr-14-103

Analyze and Report—Tread with Care

Once data are collected, the analysis and reporting must take into consideration the context and limitations of the study, including the study population, the timeline, and numerous other factors. These should be openly discussed to provide context for the study's results. For example, a cross-sectional study may be relatively easy and affordable but not rigorous enough to claim impact-level results. Similarly, although a study without a control group may find improved outcomes within a population over time, it may remain unclear whether they result from the intervention or from secular changes—for example, unanticipated policy and program changes; media influence; or economic changes with the potential to shape social attitudes, behaviors, and health outcomes.

Careful analysis and reporting is even more imperative in evaluations of structural interventions that, by nature, must take into account a web of confounding and intermediate variables. Evaluators should clearly define, through a TOC, how intermediate variables contribute to health outcomes. Measuring

synergies when an intervention has multiple components, as structural interventions often do, may reveal whether two components are working together to create bigger change than either could alone.

While a TOC can direct evaluators through a logical causal chain, it is important to leave flexibility for influential factors that emerge at a later time. Including mixed methods allows researchers to collect and analyze data beyond specific variables that are expected to change and possibly to explain why the intervention did or did not lead to the intended results. Project or program evaluations frequently make program planners, implementers, participants, and stakeholders nervous about what will be found and how the various parties will be portrayed. Although this is not unique to structural interventions, evaluations of biomedical or behavioral interventions put less emphasis on contextual factors and how they affect the intervention—and thus have fewer parties to "evaluate." With a structural intervention it is easier to point to the context and how it affects people's roles and behaviors as reasons for the intervention's success (or lack thereof).

Replicate and Generalize: Addressing the Question "So What?"

Structural interventions can take significant time and effort to complete, and an evaluation that finds positive results may lead program implementers and donors to claim success. However, it is important to ask, "So what?" Some important follow-up questions include: What are the implications of the results? Are they meaningful in the context of the broader population or the whole of the country? Are the findings generalizable to other contexts? Evaluators must look beyond their single study to a larger discussion in which program planners and decision makers, including policymakers, donors, evaluators, and other stakeholders, decide whether and how the structural intervention can be scaled up within the region or nation or adapted to other settings.

Since structural interventions rely heavily on the context within which they are implemented, it is challenging to determine which program components are generalizable and which are unique to that context. A recent systematic review of evaluations of structural interventions found that studies rarely mention generalizability in discussing results; a few papers do mention the lack of generalizability as a limitation (Iskarpatyoti, et al., 2016). Even for an established program such as the Sonagachi Project (Swendeman, et al., 2009), in which scale-up is emphasized, evaluators grappled with how to develop something that was both context-specific and generalizable and draw out the generalizable aspects to other sites. As one evaluator noted, "The program activities (both the implementation and the evaluation/research) are standardizable to a certain degree; but how do you know which pieces are standardizable?" In-depth conversations with local, national, and international stakeholders—that is, those with in-depth experience in specific settings and those with experience across various regions or countries—can help determine the generalizability of specific structural interventions.

Ethical Considerations

Ethical dilemmas and challenges are common in both the implementation and the evaluation of structural interventions. Challenges specific to evaluation of structural- interventions include the ethics of providing different services to the control and intervention groups, asking potentially sensitive questions of participants and project staff, ensuring the safety of the research team, and potentially affecting the intervention by means of the research process.

Key informants frequently mention the ethical issues around withholding the intervention (e.g., school fees, cash payments, policies around access to alcohol) from control groups. While providing "treatment" to the intervention group and not the control group is often necessary to examine how the intervention affects the risk of HIV transmission, the goods and services associated with the treatment are beneficial in and of themselves. For instance, paying school fees helps poor children attend school, regardless of whether increased school attendance influences the risk of HIV transmission. Sometimes researchers can

address this concern within the evaluation design. For example, a national government implementing a policy to eliminate school fees might take a phased approach, rolling out the policy in different parts of the country over 24 months. Researchers could capitalize on this phased rollout by identifying control and intervention groups at the outset (when the policy is being implemented in some regions but not others) but allowing the initial control areas to benefit from the policy at the end of the rollout period.

Another common challenge is navigating the relationship between evaluators and implementers, particularly when data gathered for the evaluation suggests a need to alter aspects of the intervention. Evaluators must decide whether to provide real-time feedback, potentially altering how the intervention is rolled out and thus influencing evaluation results.

In order to identify and plan for ethical considerations, it is important to engage with stakeholders from the beginning of an intervention and throughout the evaluation process, and to carefully consider the overall design and methods to be used in the evaluation. Stakeholder engagement can be time consuming, and donors must understand the need for careful planning. In some contexts, research may be necessary to illuminate the ethical and other challenges posed by an evaluation.

CONCLUSION

Although numerous challenges exist in evaluating structural interventions, current evidence indicates their promise for HIV prevention. By planning for evaluations early, working with stakeholders, determining a theory of change, carefully selecting research questions, and selecting the most appropriate research design, the HIV-prevention field can continue to determine which types of interventions will be most effective.

In addition to contributing to the body of evidence on the effectiveness of structural interventions for HIV prevention, these evaluations can serve as templates for researchers examining other types of complex interventions.

REFERENCES

Alexander, L. K., Lopes, B., Ricchetti-Masterson, K. & Yeatts, K. B. (n.d.). *Cross-sectional Studies*. Retrieved from https://sph.unc.edu/files/2015/07/nciph_ERIC8.pdf

Auerbach, J. D., Parkhurst, J. O. & Cáceres, C. F. (2011). Addressing social drivers of HIV/AIDS for the long-term response: conceptual and methodological considerations. *Global Public Health, 6* (sup3), S293–S309.

Bamberger, M., Rugh, J. & Mabry, L. (2011). *RealWorld evaluation: Working under budget, time, data, and political constraints*: Sage.

BetterEvaluation. (2014). Triangulation. Retrieved from http://betterevaluation.org/evaluation-options/triangulation

BetterEvaluation. (n.d.-a). Developmental Evaluation. Retrieved from http://betterevaluation.org/plan/approach/developmental_evaluation

BetterEvaluation. (n.d.-b). Randomized Controlled Trial (RCT). Retrieved from http://betterevaluation.org/plan/approach/rct

Blankenship, K. M., Friedman, S. R., Dworkin, S. & Mantell, J. E. (2006). Structural interventions: concepts, challenges and opportunities for research. *Journal of Urban Health, 83* (1), 59–72.

Bonell, C., Hargreaves, J., Strange, V., Pronyk, P. & Porter, J. (2006). Should structural interventions be evaluated using RCTs? The case of HIV prevention. *Social Science & Medicine, 63* (5), 1135–1142.

Campbell, M., Fitzpatrick, R., Haines, A. & Kinmonth, A. L. (2000). Framework for design and evaluation of complex interventions to improve health. *British medical journal, 321* (7262), 694.

Cook, T. D. (2002). Randomized experiments in educational policy research: A critical examination of the reasons the educational evaluation community has offered for not doing them. *Educational Evaluation and Policy Analysis, 24* (3), 175–199.

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I. & Petticrew, M. (2008). Developing and evaluating complex interventions: the new Medical Research Council guidance. *Bmj, 337*, a1655.

Davies, R. & Dart, J. (2005). The 'most significant change' (MSC) technique. *A guide to its use*.

De Silva, M. J., Breuer, E., Lee, L., Asher, L., Chowdhary, N., Lund, C., & Patel, V. (2014). Theory of Change: a theory-driven approach to enhance the Medical Research Council's framework for complex interventions. Trials, 15(1), 1.

Fredericks, K. & Carman, J. (2013). *Using Social Network Analysis in Evaluation*. Retrieved from http://www.rwjf.org/en/library/research/2013/12/using-social-network-analysis-in-evaluation.html

Gupta, G. R., Parkhurst, J. O., Ogden, J. A., Aggleton, P. & Mahal, A. (2008). Structural approaches to HIV prevention. *The Lancet, 372* (9640), 764–775.

Habicht, J.-P., Victora, C. & Vaughan, J. P. (1999). Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact. *International journal of epidemiology, 28* (1), 10–18.

Hall, H. R. & Roussel, L. A. (2012). *Evidence-based practice*: Jones & Bartlett Publishers.

Hankivsky, O. (2014). Intersectionality 101. *cal, 64* (1), 238.

Heise, L. & Watts, C. (2013). Intervening upstream: a good investment for HIV prevention.

Institute of Medicine. (2013). *Evaluation of PEPFAR*. Washington, DC: The National Academies Press.

Iskarpatyoti, B. S., Lorenzetti, L., Hart, L., Thomas, J. & Mandal, M. (2016). *Evaluations of Structural Interventions for HIV Prevention: A Systematic Review*.

Kennedy, C. (2015). *Interviewer: M. Mandal.*

Krishnan, S., Dunbar, M. S., Minnis, A. M., Medlin, C. A., Gerdts, C. E., & Padian, N. S. (2008). Poverty, gender inequities, and women's risk of human immunodeficiency virus/AIDS. Annals of the New York Academy of Sciences, 1136(1), 101-110.

Lennie, J. (2011). *The Most Significant Change technique: A manual for M&E staff and others at Equal Access* Retrieved from Equal Access Participatory Monitoring and Evaluation toolkit:
Mann, C. (2003). Observational research methods. Research design II: cohort, cross sectional, and case-control studies. *Emergency Medicine Journal, 20* (1), 54–60.

MEASURE Evaluation. (2011). *Stakeholder Engagement Tool.* Retrieved from https://www.measureevaluation.org/resources/publications/ms-11-46-e

Parkhurst, J. O. (2013). Structural drivers, interventions and approaches for prevention of sexually transmitted HIV in general populations: Definitions and an operational approach.

Pronyk, P. & Lutz, B. (2013). Policy and programme responses for addressing the structural determinants of HIV. Arlington, VA: USAID's AIDS Support and Technical Assistance Resources, AIDSTAR-One, Task Order, 1.

Sommer, M. & Parker, R. (2013). *Structural approaches in public health*: Routledge.

Swendeman, D., Basu, I., Das, S., Jana, S., & Rotheram-Borus, M. J. (2009). Empowering sex workers in India to reduce vulnerability to HIV and sexually transmitted diseases. Social Science & Medicine, 69(8), 1157-1166.

The Office of the U.S. Global AIDS Coordinator. (2014). *PEPFAR 3.0: Controlling the Epidemic: Delivering on the Promise of an AIDS-free Generation*. Retrieved from http://www.pepfar.gov/documents/organization/234744.pdf

Thomas, J. C., Curtis, S. & Smith, J. B. (2011). The broader context of implementation science. *JAIDS Journal of Acquired Immune Deficiency Syndromes, 58* (1), e19–e21.

Thomas, J. C., Reynolds, H. W., Alterescu, X., Bevc, C., & Tsegaye, A. (2016). Improving referrals and integrating family planning and HIV services through organizational network strengthening. Health policy and planning, 31(3), 302-308.

Tumlinson, K., Thomas, J. C., & Reynolds, H. W. (2015). The effect of women's property rights on HIV: a search for quantitative evidence. AIDS care, 27(1), 112-122.
USAID. (2010). *Performance Monitoring & Evaluation: Conducting Mixed-Method Evaluations.*

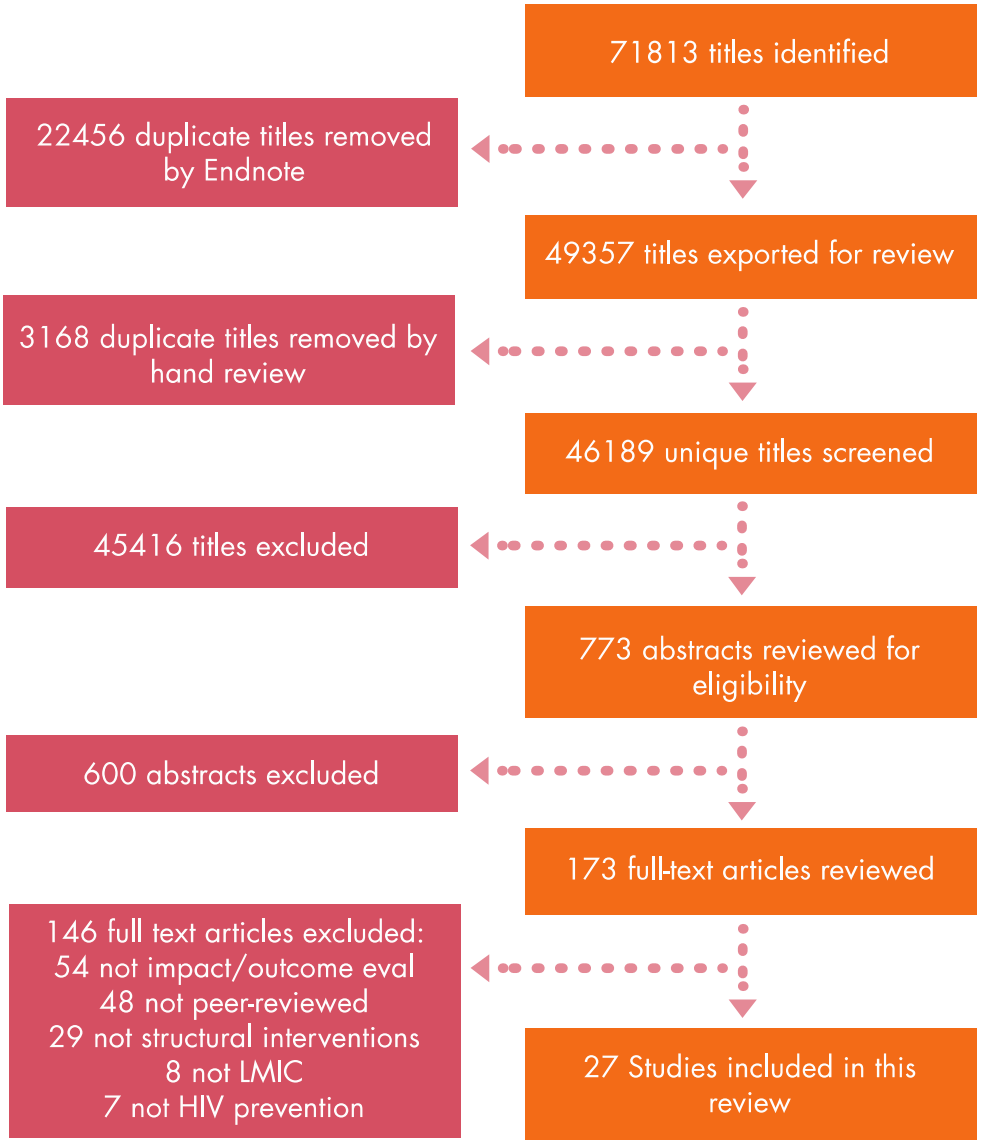
USAID. (2013). *Discussion Note: Complexity-Aware Monitoring*. Retrieved from https://usaidlearninglab.org/library/complexity-aware-monitoring-discussion-note-brief

USAID Asia. (2010). *Data Triangulation for HIV Prevention Program Evaluation in Low and Concentrated Epidemics*. Retrieved from https://www.fhi360.org/resource/data-triangulation-hiv-prevention-program-evaluation-low-and-concentrated-epidemics-0

Wilson-Grau, R. (2015). Outcome Harvesting. Retrieved from http://betterevaluation.org/plan/approach/outcome_harvesting

Wilson-Grau, R. & Britt, H. (2013). *Outcome Harvesting*. Retrieved from https://usaidlearninglab.org/sites/default/files/resource/files/Outome Harvesting Brief FINAL 2012-05-2-1.pdf

APPENDIX A. DIAGRAM OF THE SYSTEMATIC REVIEW PROCESS



APPENDIX B. BIBLIOGRAPHY OF ARTICLES INCLUDED IN EVALUATIONS OF STRUCTURAL INTERVENTIONS FOR HIV PREVENTION: A SYSTEMATIC REVIEW

1. Ashburn, K., Kerrigan, D., & Sweat, M. (2008). Micro-credit, women's groups, control of own money: HIV-related negotiation among partnered Dominican women. *AIDS and Behavior, 12*(3), 396–403. Retrieved from http://link.springer.com/article/10.1007%2Fs10461-007-9263-2.

2. Austrian, K., & Muthengi, E. N. (2014). Can economic assets increase girls' risk of sexual harassment? Evaluation results from a social, health and economic asset-building intervention for vulnerable adolescent girls in Uganda. *Children and Youth Services Review, 47*(2), 168–175. Retrieved from http://www.popcouncil.org/research/can-economic-assets-increase-girls-risk-of-sexual-harassment-evaluation-res.

3. Baird, S. J., Garfein, R. S., McIntosh, C. T., & Özler, B. (2012). Effect of a cash transfer programme for schooling on prevalence of HIV and herpes simplex type 2 in Malawi: A cluster randomised trial. *The Lancet, 379*, 1320–29. Retrieved from http://www.thelancet.com/journals/lancet/article/PIIS0140-6736(11)61709-1/abstract.

4. Baird, S., Chirwa, E., McIntosh. C., & Özler, B. (2010). The short-term impacts of a schooling conditional cash transfer program on the sexual behavior of young women. *Health Economics, 19*(S1), 55–68. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/19946887.

5. Blanchard, A. K., Mohan, H. L., Shahmanesh, M., Prakash, R., Isac, S., Ramesh, B. M., . . . Blanchard, J. F. (2013). Community mobilization, empowerment and HIV prevention among female sex workers in south India. *BMC Public Health*, *13*(234). Retrieved from http://bmcpublichealth.biomedcentral.com/articles/10.1186/1471-2458-13-234.

6. Cho, H., Hallfors, D. D., Mbai, I. I., Itindi, J., Milimo, B. W., Halpern, C. T., & Iritani, B. J. (2011). Keeping adolescent orphans in school to prevent human immunodeficiency virus infection: Evidence from a randomized controlled trial in Kenya. *Journal of Adolescent Health, 48*(5), 523–526. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/21501814.

7. Cluver, L., Boyes, M., Orkin, M., Pantelic, M., Molwena, T., & Sherr, L. (2013). Child-focused state cash transfers and adolescent risk of HIV infection in South Africa: A propensity-score-matched case-control study. *The Lancet Global Health, 1*(6), 362–370. Retrieved from http://www.thelancet.com/journals/langlo/article/PIIS2214-109X(13)70115-3/abstract.

8. De Walque, D., Dow, W. H., Nathan, R., Abdul, R., Abilahi, F., Gong, E., . . . Medlin, C. A. (2012). Incentivising safe sex: A randomised trial of conditional cash transfers for HIV and sexually transmitted infection prevention in rural Tanzania. *BMJ Open, 2*, 000747. Retrieved from http://bmjopen.bmj.com/content/2/1/e000747.abstract.

9. Dunbar, M. S., Dufour, M. K., Lambdin, B., Mudekunye-Mahaka, I., Nhamo, D., & Padian, N. S. (2014). The SHAZ! project: Results from a pilot randomized trial of a structural intervention to prevent HIV among adolescent women in Zimbabwe. *PLoS One*, *9*(11), 113621. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/25415455.

10. Euser, S. M., Souverein, D., Gowda, P. R. N., Gowda, C. S., Grootendorst, D., Ramaiah, R., . . . Den Boer, J. W. (2012). Pragati: An empowerment programme for female sex workers in Bangalore, India. *Global Health Action, 5*, 19279. Retrieved from http://www.globalhealthaction.net/index.php/gha/article/view/19279.

11. Hallfors, D. D., Cho, H., Mbai, I., Milimo, B., & Itindi, J. (2012). Process and outcome evaluation of a community intervention for orphan adolescents in western Kenya. *Journal of Community Health, 37*(5), 1101–1109. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3392443/.

12. Hallfors, D., Cho, H., Rusakaniko, S., Iritani, B., Mapfumo, J., & Halpern, C. (2011). Supporting adolescent orphan girls to stay in school as HIV risk prevention: Evidence from a randomized controlled trial in Zimbabwe. *American Journal of Public Health, 101*(6), 1082–1088. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/21493943.

13. Kim, J., Ferrari, G., Abramsky, T., Watts, C., Hargreaves, J., Morison, L., . . . Pronyk, P. (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), 824–832. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/20072767.

14. Kohler H., & Thornton R. L. (2011). Conditional cash transfers and HIV/AIDS prevention: Unconditionally promising? *The World Bank Economic Review, 26*(2), 165–190. Retrieved from. http://elibrary.worldbank.org/doi/abs/10.1093/wber/lhr041.

15. Magnani, R. J., McCann, H. G., Hotchkiss, D. R., & Florence, C. S. (1998). The effects of monetized food aid on reproductive behavior in rural Honduras. *Population Research and Policy Review, 17*(4), 305–328. Retrieved from http://link.springer.com/article/10.1023/A:1005926905494.

16. Odek, W. O., Busza, J., Morris, C. N., Cleland, J., Ngugi, E. N., & Ferguson, A. G. (2009). Effects of micro-enterprise services on HIV risk behaviour among female sex workers in Kenya's urban slums. *AIDS and Behavior, 13*(3), 449–461. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/18998204.

17. Pronyk, P. M., Kim, J. C., Abramsky, T., Phetla, G., Hargreaves, J. R., Morison, L. A., . . . Porter, J. D. (2008). A combined microfinance and training intervention can reduce HIV risk behaviour in young female participants. *Aids, 22*(13), 1659–1665. Retrieved from http://www.sef.co.za/files/08%20-%20Pronyk%20IMAGE%20on%20HIV%20AIDS%202008.pdf.

18. Pronyk, P. M., Hargreaves, J. R., Kim, J. C., Morison, L. A., Phetla, G., Watts, C., . . . Porter, J. D. (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), 1973–1983. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/17141704.

19. Rosenberg, M., Pettifor, A., Thirumurthy, H., Halpern, C. T., & Handa, S. (2014). The impact of a national poverty reduction program on the characteristics of sex partners among Kenyan adolescents. *AIDS and Behavior, 18*(2), 311–316. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/23612944.

20. Rosenberg, M. S., Seavey, B. K., Jules, R., & Kershaw, T. S. (2011). The role of a microfinance program on HIV risk behavior among Haitian women. *AIDS and Behavior, 15*(5), 911–918. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/21153762.

21. Sherer, R. D., Bronson, J. D., Teter, C. J., & Wykoff, R. F. (2004). Microeconomic loans and health education to families in impoverished communities: Implications for the HIV pandemic. *Journal of the International Association of Physicians in AIDS Care, 3*(4), 110–114. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/15768731.

22. Souverein, D., Euser, S. M., Ramaiah, R., Gowda, P. R. N., Gowda C. S., Grootendorst, D. C., . . . Den Boer, J. W. (2013). Reduction in STIs in an empowerment intervention programme for female sex workers in Bangalore, India: The Pragati programme. *Global Health Action, 6*, 22943. Retrieved from http://www.globalhealthaction.net/index.php/gha/article/view/22943.

23. Spielberg, F., Crookston, B. T., Chanani, S., Kim, J., Kline, S., & Gray, B. L. (2013). Leveraging microfinance to impact HIV and financial behaviors among adolescents and their mothers in West Bengal: A cluster randomized trial. *International Journal of Adolescent Medicine and Health, 25*(2), 157–166. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/23324373.

24. Ssewamala, F. M., Han, C., Neilands, T. B., Ismayilova, L., & Sperber, E. (2010). Effect of economic assets on sexual risk-taking intentions among orphaned adolescents in Uganda. *American Journal of Public Health, 100*(3), 483–488. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2820050/.

25. Ssewamala, F. M., Ismayilova, L., McKay, M., Sperber, E., Bannon, W., & Alicea, S. (2010). Gender and the effects of an economic empowerment program on attitudes toward sexual risk-taking among AIDS-orphaned adolescent youth in Uganda. *Journal of Adolescent Health, 46*(4), 372–378. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2844862/.

26. Ssewamala, F. M., & Ismayilova, L. (2009). Integrating children's savings accounts in the care and support of orphaned adolescents in rural Uganda. *The Social Service Review, 83*(3), 453–472. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2863345/.

27. Swendeman, D., Basu, I., Das, S., Jana, S., & Rotheram-Borus, M. J. (2009). Empowering sex workers in India to reduce vulnerability to HIV and sexually transmitted diseases. *Social Science & Medicine, 69*(8), 1157–1166. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/19716639.

APPENDIX C. GRAY LITERATURE SEARCH

A gray literature search for relevant resources was conducted. The following databases, organizational websites and conference abstracts were searched:

Databases of Working Papers

Social Science Research Network

Project/Organization Websites

STRIVE

AIDSTAR-One

AIDSTAR-Two

AIDSFree

Population Council

The Palladium Group

Johns Hopkins School of Public Health

Pacific Institute for Research and Evaluation

International Center for Research on Women

FHI360

ASPIRES

Conference and Other Abstracts

APHA

International AIDS Conference

American Evaluation Conference

Database of Abstracts of Reviews of Effectiveness (DARE)

Other Sources for Ongoing Research

NIH Research Portfolio Online Reporting Tool

APPENDIX D. KEY INFORMANTS

Caitlin Kennedy, Associate Professor and Director of Social and Behavioral Interventions Program, Department of International Health at Johns Hopkins Bloomberg School of Public Health

Clare Barrington, Associate Professor, Department of Health Behavior at the University of North Carolina Gillings School of Global Public Health

Dallas Swendeman, Assistant Professor, Department of Psychiatry and Biobehavioral Sciences at the University of California, Los Angeles, David Geffen School of Medicine; and Affiliated Faculty, Department of Epidemiology at the University of California, Los Angeles Fielding School of Public Health

Denise Halfors, Senior Research Scientist, Pacific Institute for Research and Evaluation

Hyunson Cho, Research Scientist, Pacific Institute for Research and Evaluation

Jhumka Gupta, Assistant Professor, Department of Global and Community Health at George Mason University

Kelly Kilburn, Research Assistant, Department of Public Policy at the University of North Carolina

Kim Blankenship, Professor and Chair, Department of Sociology; and Director of the Center on Health, Risk and Society at American University

Ravi Verma, Director of the Asia Regional Office, International Center for Research on Women

Diana Kerrigan, Associate Professor, Department of Health, Behavior and Society; and Project Director of the Research to Practice Project at Johns Hopkins Bloomberg School of Public Health

Susan Sherman, Professor, Departments of Epidemiology; and Director, Baltimore Center for AIDS Research at Johns Hopkins Bloomberg School of Public Health.

MEASURE Evaluation
Carolina Population Center
University of North Carolina at Chapel Hill
400 Meadowmont Village Circle, 3rd Floor
Chapel Hill, North Carolina 27517
Phone: +1 919-445-9350 | Fax: +1 919-445-9353
measure@unc.edu
www.measureevaluation.org

This publication was produced with the support of the United States Agency for International Development (USAID) under the terms of MEASURE Evaluation cooperative agreement AID-OAA-L-14-00004. MEASURE Evaluation is implemented by the Carolina Population Center, University of North Carolina at Chapel Hill in partnership with ICF International; John Snow, Inc.; Management Sciences for Health; Palladium; and Tulane University. Views expressed are not necessarily those of USAID or the United States government. TR-16-139

USAID FROM THE AMERICAN PEOPLE

PEPFAR

MEASURE Evaluation