Evaluations of Structural Interventions for HIV Prevention: A Review of Approaches and Methods

Brittany S. Iskarpatyoti[1] · Jill Lebov[2] · Lauren Hart[1] · Jim Thomas[1] · Mahua Mandal[1]

Published online: 22 December 2017

Abstract

Structural interventions alter the social, economic, legal, political, and built environments that underlie processes affecting population health. We conducted a systematic review of evaluations of structural interventions for HIV prevention in low- and middle-income countries (LMICs) to better understand methodological and other challenges and identify effective evaluation strategies. We included 27 peer-reviewed articles on interventions related to economic empowerment, education, and substance abuse in LMICs. Twenty-one evaluations included clearly articulated theories of change (TOCs); 14 of these assessed the TOC by measuring intermediary variables in the causal pathway between the intervention and HIV outcomes. Although structural interventions address complex interactions, no evaluation included methods designed to evaluate complex systems. To strengthen evaluations of structural interventions, we recommend clearly articulating a TOC and measuring intermediate variables between the predictor and outcome. We additionally recommend adapting study designs and analytic methods outside traditional epidemiology to better capture complex results, influences external to the intervention, and unintended consequences.

Keywords Structural interventions · Evaluation · HIV · Prevention · Low- and middle-income countries

Introduction

Structural interventions refer to public health interventions that alter the social, economic, legal, political, and built environment that shapes health processes and outcomes [1, 2]. While they have a long history of implementation and have been considered successful in sectors such as water and sanitation (e.g., water purification and latrine construction) [3, 4], structural interventions have only recently drawn attention in the HIV-prevention field. The World Health Organization's Global Health Sector Strategy for HIV/AIDS 2011–2015 included for the first time the removal of structural barriers as one of its four strategic directions to achieve universal access to HIV prevention, diagnosis, treatment, and care services [5]. Similarly, the Institute of Medicine's (IOM's) 2013 evaluation of the United States President's Emergency Plan for AIDS Relief (PEPFAR) observed that structural interventions constitute the smallest proportion of PEPFAR's response thus far. The IOM recommended a stronger emphasis on prevention, using a balanced selection of biomedical, behavioral, and structural interventions [6]. Development professionals have suggested that insufficient attention to structural factors has inhibited HIV prevention efforts [7].

In response to this emerging emphasis, several large-scale structural interventions have been implemented in low- and middle-income countries (LMICs) with high HIV prevalence. For example, because higher levels of education have been shown to reduce HIV risk among young women and girls [8–10], interventions in South Africa and Kenya have provided female adolescents and youth with cash transfers conditioned on school attendance [11–13]. Other types of structural interventions include microenterprise interventions that encourage financial planning and distribution of

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10461-017-1997-x) contains supplementary material, which is available to authorized users.

✉ Mahua Mandal
mmandal@email.unc.edu

[1] MEASURE Evaluation, Carolina Population Center, University of North Carolina, CB 3446, Chapel Hill, NC 27599, USA

[2] RTI International, Durham, NC, USA

small loans to help reduce HIV risk behavior among young females [14–16].

Evaluations of structural interventions face several methodological and implementation challenges. Structural interventions are affected by and frequently implemented at multiple levels (e.g., individual, community, and policy levels) [17], making it difficult to employ randomized control trials (RCTs), the traditional "gold standard" for biomedical and public health evaluations [18]. Random assignment of groups may not be feasible or ethical [19]. Without random assignment it is difficult to rule out selection bias, where individuals exposed to the intervention are different from unexposed individuals; this obscures the intervention's effect on health outcomes. Relatedly, structural interventions, which are often multisectoral and complex, are typically comprised of many parts that interact with each other as well as with the built and social environment [20]. This complexity has implications for measuring the potentially large number of factors that must be considered in designing and analyzing evaluation studies. Structural interventions also aim to influence factors that are "upstream," or distal (e.g., poverty, geographic location) 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 [21].

Furthermore, implementation of structural interventions may be nonlinear, iterative, and adaptive [22]; thus, using conventional methods, such as testing discrete hypotheses and measuring a predetermined set of intermediary and outcome variables, may not be suitable when evaluating structural interventions. Finally, the contextual nature of structural interventions means that factors such as economic barriers, political and legal constraints, cultural norms, and shifting power dynamics influence how an intervention is implemented by intervention staff and resonates with local communities [22]. The variability in contextual factors across settings often limits the degree to which evaluation results from one context would apply to another context.

To better understand these challenges and identify effective strategies for evaluating structural interventions, we sought to [1] review the range and rigor of approaches and methods used in evaluations of structural interventions to prevent HIV and [2] provide recommendations for improving future evaluations of structural interventions for HIV prevention.

Methods

We conducted a systematic review of methods used to eval-uate the outcomes and impacts of structural interventions for HIV prevention, focusing on economic empowerment, access to formal and informal education, and reduction of substance use, in LMICs.

Search Strategy and Selection Criteria

Our review methods followed an adapted version of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. We searched for peer-reviewed articles in PubMed, PsycINFO, Scopus, POPLINE, EconLit, Social Science Citation Index, and Global Health. We limited our search to articles published between January 1, 1990 and December 31, 2015 to focus on recent methods and approaches used to evaluate structural interventions for HIV prevention. Search terms included Medical Subjects Heading (MeSH) and other associated terms for HIV as well as terms related to economic empowerment; formative, informal, and formal education (e.g., "primary school"); or substance use prevention (see supplementary files for complete search string). To be included in this review the evaluation must have (a) been conducted in an LMIC, as defined by the World Bank [23]; (b) described an impact or outcome evaluation; and (c) assessed outcomes that were explicitly related to HIV prevention (e.g., condom use, reducing multiple partners, etc.). We included evaluation studies that used quantitative-only or both quantitative and qualitative methods, an approach that permitted inclusion of evaluation designs ranging from RCTs to quasi- and non-experimental evaluations. We excluded articles that described process evaluations or monitoring data only. Interventions that addressed family planning or pregnancy prevention outcomes without also including HIV prevention as an objective were excluded from our review. Biomedical studies, nonhuman research, and studies focused solely on individual counseling or education interventions to increase HIV-prevention awareness or sexual health knowledge were also excluded.

Article Screening and Selection

Article citations were uploaded into the reference management software EndNote X7, and duplicate articles were automatically removed using the de-duplication feature. We then exported the title, author, year of publication, journal, and name of database into a Microsoft Excel spreadsheet for title and abstract review. Two of four authors (BS, MM, JL, LH) independently screened the titles to exclude studies that were not relevant. Discrepancies were resolved through discussions between the two original reviewers or, if necessary, referral to a third reviewer. For potentially relevant titles, reviewers screened the abstracts. If the abstract did not provide adequate information to determine whether the article should be included, we reviewed the full text of the article to determine its relevance (see Fig. 1). In cases

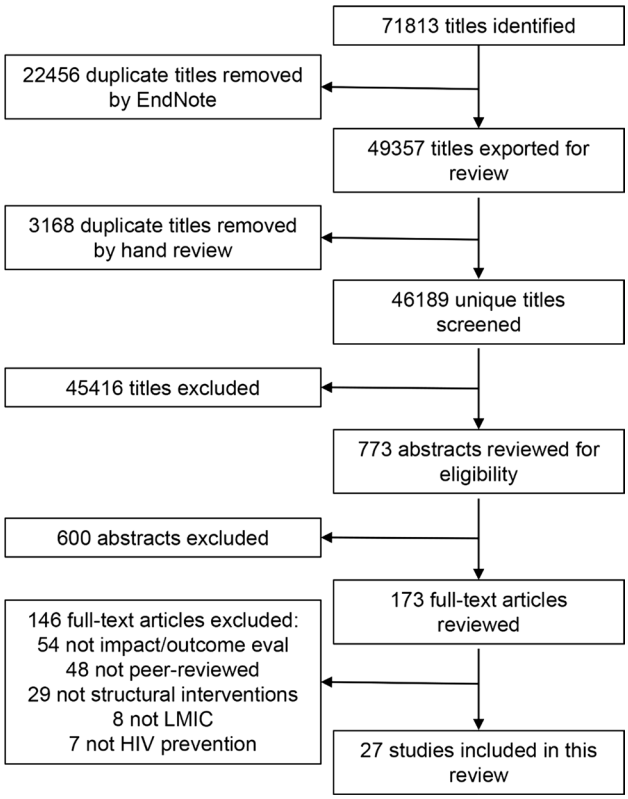

Fig. 1 Literature review process for review of evaluations of structural interventions related to HIV prevention in low- and middle-income countries

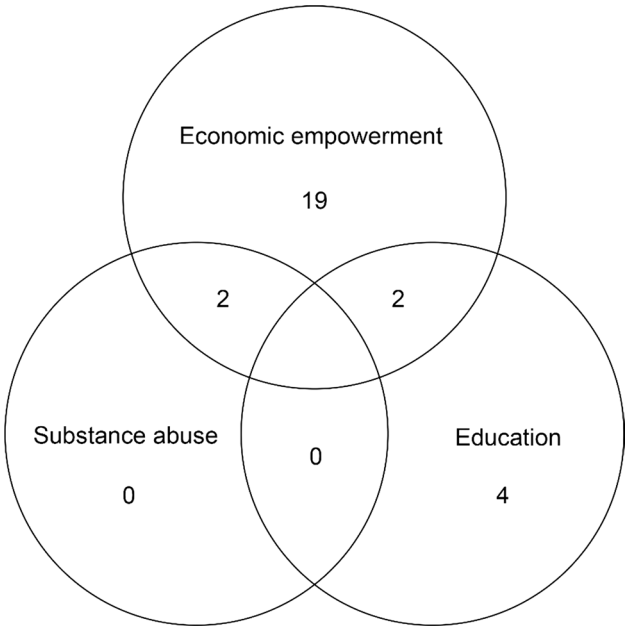

Fig. 2 Diagram of included evaluations by type of structural intervention

where multiple relevant articles evaluated the same intervention, we included each article that used a distinct evaluation approach.

Data Abstraction and Analysis

For all relevant peer-reviewed articles, we abstracted data for the following: [1] information on the intervention being evaluated, including type of structural intervention, target populations, and start and end date of the intervention; [2] components of the theory of change (TOC) or causal pathway framework including predictor variables, intermediate variables, and outcomes of interest; [3] components of the evaluation, including type of evaluation (outcome or impact), type of data collected, timeline for data collection, and statistical methods used to analyze data; [4] study limitations; and [5] reported generalizability of the findings and scalability of the intervention.

To systematically compare the studies reported in the articles, the rigor of each evaluation was assessed using a checklist informed by published systematic reviews examining the quality of studies [24, 25]. The checklist used 13 items to evaluate the study's methodological quality: inclusion of a theoretical framework that guided the intervention or study; use of mixed methods (included in the article or referenced); randomized assignment of participants; prolonged engagement in the study (equal to or greater than 18 months); justification of sample size; inclusion of a cohort; inclusion of a control or comparison group; a comparison group that, at baseline, is equivalent to the study group on outcome variables; a comparison group that is equivalent to the study group on sociodemographic characteristics; availability of pre and post data; a follow-up rate of 80 percent or higher; statistical significance testing; and reporting of intervention implementation details to facilitate replicability. Studies received one point for fulfillment of each criterion, and the range of possible points was 0–13. Points were summed, and a rigor score was applied using the following rubric: 0–4 low; 5–9 medium; 10–13 high.

Results

The search yielded 27 articles, describing evaluations of 20 interventions. Of the 27 articles, 23 were evaluations of economic empowerment interventions, six were of education interventions, and two included a substance abuse intervention with some overlap in intervention approaches (see Fig. 2). Most evaluations came from interventions in sub-Saharan Africa (n = 18). Other evaluations were implemented in South Asia (n = 5) and Latin America and the Caribbean (n = 3); one evaluation had a global scope.

Theory of Change

Over three-quarters (n = 21) of articles included a description of a TOC that guided their intervention or evaluation. Among evaluations that did not include a TOC (n = 6), three proposed causal pathways post hoc [26–28]. Several studies developed TOCs specifically for the intervention (n = 9). For example, the Avahan Initiative developed an Integrated Empowerment Framework for female sex workers that describes three components of power (power within, power with others, power over resources) which lead to social and personal transformation for better health behaviors and positive health outcomes [29]. Other studies were guided by established models or theories, including the Freierian and Chen model [30], Asset Theory [14, 31–33], the Social Development Model [34–36], empowerment theories [37–39], and Connell's theory of gender and power [15].

The SHAZ! Project developed a theoretical framework that described the pathway between their intervention and the intended HIV outcomes: access to economic opportunities; HIV education; and HIV prevention, care, and treatment. Participation in the intervention was hypothesized to contribute to economic and social empowerment, improve relationship power and reduce sexual risk behaviors and gender-based violence (GBV). This, in turn, would lead to a reduction in HIV and other sexually transmitted infections. Dunbar et al., mirrored this framework in the design of their evaluation, measuring intermediate variables for economic and social factors, relationship power, and sexual risk behaviors. They additionally included measures of biological factors to measure impact of the intervention on HIV infection [40]. This framework was published alongside the results of the evaluation to illustrate the possible causal links among factors.

Two-thirds of the evaluations that included a TOC measured intermediary TOC components including various dimensions of empowerment, gender norms, and attitudes and behaviors. Seven articles that described a TOC measured the association of the intervention activities with HIV outcomes but did not measure the intermediary variables in the causal pathway. Two articles referenced other articles from the same study that measured intermediary variables [14, 38].

Study Design

RCTs were the most common quantitative study design (n = 16), with the remaining evaluations using quasi-experimental (n = 7) and non-experimental designs (n = 4). The majority (n = 25) of evaluations used quantitative methods exclusively; three of these articles referenced publications from the same study reporting qualitative results [29, 39, 40]. Two articles reported both quantitative and qualitative methods and results within the same article [14, 36]; however, one collected qualitative measures for process monitoring only [36]. One article incorporated qualitative methods to measure the study outcome [40]. Pronyk et al., used thematic content analysis of observations of loan center meetings, focus group discussions, key informant interviews, and diaries of training facilitators to provide context and understand why the IMAGE intervention led to certain expected outcomes and not others in South Africa [14].

All articles provided basic sampling information such as sample size and sampling methods (i.e., randomized, matched, convenience). Eleven studies included additional information regarding sample size justification such as power calculations. Interventions commonly targeted specific populations including women or girls (n = 17), adolescents or youth (n = 15), orphans and vulnerable children (n = 8), female sex workers (n = 5), and adult couples (n = 2). Three-quarters (n = 20) of the articles included a control or comparison group, of which 90 percent (n = 18) were equivalent on socio-demographic characteristics and 65 percent (n = 13) were equivalent on outcome measures, compared to the intervention group at baseline (see Table 1).

The length of interventions varied widely from 10 months to 10 years. Study timelines generally mirrored intervention timelines (baseline measures collected at the start of interventions; end line measures collected at the end of interventions), with the exception of three cross-sectional studies [28–30] that measured outcomes at one point in time. While most studies included multiple data collection time points (n = 24), no study conducted ex-post data collection and analysis to determine long-term effects of the intervention. Although the timelines for most studies were short (mean 18 months; median 12 months), many studies experienced high levels of attrition. Thirty-seven percent of studies (n = 10) reported participant follow-up rates of less than 80 percent between baseline and end line.

All studies included outcome-level and seven studies included impact-level measures. Study designs mapped closely to the type of evaluation; that is, evaluations that included impact measures more often used an RCT design (see Fig. 3).

The most frequent outcomes measured were sexual and HIV and AIDS knowledge (n = 11), attitudes (n = 9), and behaviors (n = 24). Couple communication and condom negotiation (n = 8), gender based violence (n = 6) and gender attitudes and norms (n = 6) were also included as outcome variables. Articles that measured impact on HIV infection collected biomarker data from rapid oral and/or blood tests to determine HIV incidence rates (n = 7). Two articles included measures of STI incidence, but did not test specifically for HIV and were not considered HIV impact evaluations. Due to relatively low incidence, most evaluations that included HIV impact measures were not adequately powered

Table 1 Rigor scores for included evaluations

First author	Year	Type of structural intervention	Type of evaluation	Study design	Theoretical framework/TOC	Mixed methods	Random assignment	Prolonged engagement in study setting (≥ 18 months)	Sample size justified	Cohort	Control or comparison group (quantity)	Comparison groups equivalent on sociodemographics	Comparison groups equivalent at baseline on outcome measure	Pre-/post-data	Follow-up rate of 80% or more	Statistical significance testing	Detail to facilitate replication	Points	Rigor: 0–4 weak; 5–9 moderate; 10–13 strong
S. Baird	2012	Economic empowerment; education	Impact	RCT	1	0	1	1	1	1	1	1	1	1	1	1	1	12	High
D. Hallfors	2012	Education	Outcome	RCT	1	1	1	1	0	1	1	1	1	1	1	1	1	12	High
P. M. Pronyk	2008	Economic empowerment	Impact and outcome	RCT	1	1	1	1	0	0	1	1	1	1	1	1	1	11	High
M. S. Dunbar	2014	Economic Empowerment	Impact and outcome	RCT	1	1	1	1	0	1	1	1	1	1	0	1	1	11	High
P. M. Pronyk	2006	Economic empowerment	Impact and outcome	RCT	1	0	1	1	1	1	1	1	0	1	1	1	1	11	High
M. Rosenberg	2014	Economic empowerment	Outcome	RCT	1	0	1	1	1	1	1	1	1	1	0	1	1	11	High
F. M. Ssewamala	2009	Economic empowerment; education	Outcome	RCT	1	0	1	0	1	1	1	1	1	1	1	1	1	11	High
H. Cho	2011	Education	Outcome	RCT	1	0	1	0	0	1	1	1	1	1	1	1	1	10	High
D. Hallfors	2011	Education	Outcome	RCT	1	0	1	1	0	1	1	1	0	1	1	1	1	10	High

Table 1 (continued)

First author	Year	Type of structural intervention	Type of evaluation	Study design	Theoretical framework/TOC	Mixed methods	Random assignment	Prolonged engagement in study setting (≥ 18 months)	Sample size justified	Cohort	Control or comparison group (quantity)	Comparison groups equivalent on socio-demographics	Comparison groups equivalent at baseline on outcome measure	Pre-/post-data	Follow-up rate of 80% or more	Statistical significance testing	Detail to facilitate replication	Points	Rigor: 0–4 weak; 5–9 moderate; 10–13 strong
D. Swendeman	2009	Economic empowerment	Outcome	Quasi-experimental	1	1	0	0	0	0	1	1	1	1	1	1	1	9	Medium
D. de Walque	2012	Economic empowerment	Impact and outcome	RCT	1	0	1	0	1	0	1	1	1	1	0	1	1	9	Medium
F. M. Ssewamala	2010	Economic Empowerment	Outcome	RCT	1	0	1	0	0	0	1	1	1	1	1	1	1	9	Medium
F. M. Ssewamala	2010	Economic empowerment	Outcome	RCT	1	0	1	0	1	0	1	1	1	1	0	1	1	9	Medium
L. Cluver	2013	Economic empowerment	Outcome	Quasi-experimental	0	0	0	0	1	0	1	1	1	1	1	1	1	8	Medium
S. Baird	2010	Education	Outcome	RCT	0	0	1	0	1	0	1	1	0	1	1	1	1	8	Medium
J. Kim	2009	Economic empowerment	Outcome	RCT	1	0	1	1	0	0	1	1	0	0	1	1	1	8	Medium
F. Spielberg	2013	Economic empowerment	Outcome	RCT	0	0	1	0	1	0	1	1	1	1	1	1	0	8	Medium
K. Austrian	2014	Economic empowerment	Outcome	Quasi-experimental	1	0	0	0	0	1	1	0	0	1	1	1	1	7	Medium
H-P. Kohler	2012	Economic empowerment	Outcome	RCT	0	0	1	0	0	1	1	0	0	1	1	1	1	7	Medium

Table 1 (continued)

First author	Year	Type of structural intervention	Type of evaluation	Study design	Theoretical framework/TOC	Mixed methods	Random assignment	Prolonged engagement in study setting (≥ 18 months)	Sample size justified	Cohort	Control or comparison group (quantity)	Comparison groups equivalent on sociodemographics	Comparison groups equivalent at baseline on outcome measure	Pre-/post-data	Follow-up rate of 80% or more	Statistical significance testing	Detail to facilitate replication	Points	Rigor: 0–4 weak; 5–9 moderate; 10–13 strong
W. O. Odek	2009	Economic empowerment	Outcome	Quasi-experimental	1	0	0	1	0	1	0	0	0	1	1	1	0	6	Medium
A. K. Blanchard	2013	Economic empowerment	Outcome	Non-experimental	1	1	0	0	1	0	0	0	0	0	0	1	1	5	Medium
S. Euser	2012	Economic empowerment; substance abuse	Outcome	Non-experimental	1	0	0	1	0	1	0	0	0	0	0	1	1	5	Medium
D. Souverein	2013	Economic empowerment; substance abuse	Impact and outcome	Quasi-experimental	1	0	0	1	0	0	0	0	0	1	0	1	1	5	Medium
M. S. Rosenberg	2011	Economic empowerment	Outcome	Non-experimental	1	0	0	0	1	0	0	0	0	0	0	1	1	4	Low
R. J. Magnani	1998	Economic empowerment	Outcome	Quasi-experimental	0	0	0	0	0	0	1	1	0	0	0	1	1	4	Low
R. D. Sherer	2004	Economic empowerment	Outcome	Quasi-experimental	0	0	0	0	0	1	0	0	0	1	1	0	1	4	Low
K. Ashburn	2008	Economic empowerment	Outcome	Non-experimental	1	0	0	0	0	0	0	0	0	0	0	1	0	2	Low

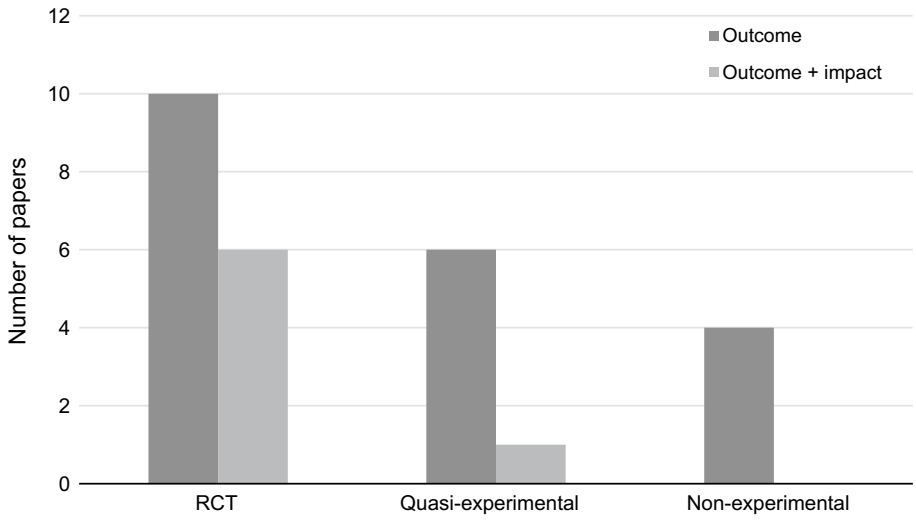

Fig. 3 Study design by type of evaluation

to detect significant results (n = 6). One study was powered to detect differences in HIV incidence between experimental and control groups [41].

Seventy percent (n = 19) of the evaluations included multivariable analyses. Of these, three-quarters (n = 14) clearly drew from the TOC in their analysis and modeling.

After aggregating the above factors related to evaluation design, implementation, and analysis of results, most studies were high (n = 9) or medium (n = 14) rigor (see Table 1). RCTs were generally rated as more rigorous than quasi-experimental studies. Many quasi-experimental studies did not provide adequate information about evaluation components, and therefore did not qualify as highly rigorous. Non-experimental designs were classified as low rigor.

Limitations, Generalizability, and Scalability

The major limitations described in these studies were attrition, short study timelines, self-selection bias, limited sample sizes, self-reported outcomes, and inability to show temporality. Nearly all studies (n = 26) included mention of limitations related to the study design, evaluation implementation, and analysis and reporting of data. While study designs are often limited by resource availability and funding timelines, many study designs included in this review were limited by the nature of the interventions. Interventions may rely on participants actively choosing to join the intervention. Among these types of interventions, self-selection bias was cited as a challenge in evaluation (n = 5). For example, empowered, lower-risk female sex workers may have been more likely to participate in a sexual risk behavior reduction intervention compared to their higher-risk, less-empowered peers [29].

In some studies (n = 11), sample sizes were not adequately powered to detect changes in specified outcome or impact indicators because of limited resources and time, unexpected barriers during recruitment, or interventions designed for hidden or smaller populations. Analyses of outcome indicators were also limited by the type of data collected; for example, many studies relied on self-reported attitude and behavioral data, which may be biased (n = 12). Short timelines that reduced exposure to an intervention limited many studies (n = 16), potentially causing them to underestimate positive outcomes that may take longer to present, or overestimate outcomes that may not be sustained. Unclear causal pathways linking interventions to outcomes was a limitation among both cross-sectional studies and those without a clear TOC (n = 12).

Replicability and generalizability were mentioned in just over half (n = 14) of the included studies, and often as a limitation or opportunity for further study (n = 9). Researchers expressed that findings could only be generalized to or replicable in similar situations, or within the same political borders. For example, Ashburn et al., mentioned that the results of their study were "generalizable only to partnered women participating in women's groups within this or similar settings" [30]. One article specifically discussed the scalability of child-focused grants within South Africa: "according to our findings, full coverage of child-focused cash transfers for the country's 2.76 million girls aged 12–18 could prevent roughly 77,000 new incidences of transactional sex each year." The authors noted, however, that study results were not generalizable outside South Africa [11].

Three-quarters of the studies (n = 20) drew conclusions that matched the results of the study. That is, the manuscripts adequately acknowledged the scope, size, and limitations of the study with regard to the observed measures of effect. For example, De Walque et al., concluded from a large (n = 5,370) 12-month RCT of conditional cash transfers in ten villages in Tanzania that "while these study results are important in showing that the idea of using financial incentives can be a useful tool for preventing HIV and STI

transmission, it remains an initial study on a limited scale" [42]. Seven studies (26 percent) drew conclusions that did not consider the scope, size, or limitations of the study, and often overstated the impact of the observed measures of effect. This was particularly prevalent among cross-sectional studies and studies that were conducted over 1 year or less.

Discussion

Our systematic review found that evaluations of structural interventions use a wide range of study designs, sample sizes, outcome measurement tools, and timelines. Most evaluations used RCTs, and a few RCTs supplemented their studies with qualitative methods. No article reported using methods and tools, such as agent-based modeling, designed to evaluate complex systems that have multiple interacting components. Agent-based modeling, a method within systems science, uses computer simulations to model the actions and interactions of individuals, or collective entities such as organizations, to examine their effects on a system.

This review also found that quasi-experimental studies show promise for rigorous evaluations when other rigor criteria, such as a clearly articulated TOC and measurement of multiple factors along the TOC, are considered in the design, implementation, and reporting of results.

Theory of Change

Because structural interventions address factors upstream from health outcomes, and operate through indirect pathways, it is essential to clearly identify these pathways through a well-articulated TOC. Evaluations that did not include a TOC may have used causal pathways to guide their studies, but did not articulate them. There is a need to develop and publish TOCs alongside the results of an evaluation to better illustrate how the intervention influences proximal factors that impact health outcomes. TOCs should include common cultural, gender, and generational norms with an eye towards the intersectionality of factors [43–48]. Measurement of intermediate variables within the TOC would provide evidence in support, or not, of the theory or parts of the theory. Additionally, measurement of outcomes of interest may be too difficult, time-consuming, or costly for researchers to obtain a sample size that provides adequate statistical power [1]. Including measures along the causal pathway can show progress towards these outcomes, even in instances where impact cannot be measured.

Study Design

While RCTs are considered the gold standard for individual-level public health evaluations, some experts argue other designs may be more appropriate for evaluating multilevel and community-level interventions [17, 19, 49]. In complex interventions randomization may not be possible, such as in the case of South Africa's ongoing cash transfer intervention [11]; ethical, such as when poverty reduction interventions cannot be withheld from a control group; or feasible, such as in interventions geared towards female sex workers where members of the target population are difficult to define [19]. An assessment of the rigor, or methodological quality, of a study must consider that study's contextual constraints [50]. Nonrandom self-selection into intervention groups has the potential to introduce bias; however, this should be weighed against other criteria, such as the acceptability and sustainability of the study design.

Structural interventions include several pathways to an outcome, and it is challenging to define a comparison or rigorous counterfactual in which all factors except the intervention are the same [18]. One-quarter of structural interventions included in this review did not include a comparison group, limiting their rigor and ability to claim that the intervention influenced HIV prevention outcomes. Researchers have argued that including qualitative methods in nonrandomized designs can address some of the challenges in identifying an appropriate counterfactual [50, 51].

Complexity-aware designs allow for the study of multifaceted and dynamic processes that may have emergent outcomes, and interactions between community- and intervention-level factors; they also depart from traditional epidemiological methods by accommodating study characteristics such as the lack of counterfactuals, bidirectional effects, feedback loops, and unpredictability [52].

Evaluations of structural interventions provide unique opportunities to apply methods such as agent-based modeling, synthetic comparisons, network analysis, and other methods that accommodate complexity. However, no single method captures all factors of and perspectives on a complex system, and using mixed method evaluations can help fill gaps in addressing evaluation questions. Qualitative data can inform how, where, when and from whom data should be collected; encourage buy-in from stakeholders; elucidate the context in which interventions take place; help interpret results; and provide insight into unanticipated results [50, 53]. Including qualitative methods in evaluations of structural interventions has the potential to improve the rigor of the study design and provide context for conclusions.

Generalizability and Scalability

While it is presumed that, due to their complexity, structural interventions have limited generalizability, it is possible to replicate the process of the intervention [17]. Including a TOC with clearly defined intermediary and intersecting components can help determine the extent to which a

structural intervention process is replicable in other contexts. For small studies with evidence of effectiveness, additional studies can assess scalability of structural interventions. Within our review, only one article discussed the potential of scale-up at a national level. Future evaluations should examine the sustainability of structural interventions, given that most in this review were evaluated within relatively short timelines.

Review Limitations

This review has several limitations. Our review included articles published in English only, which may have excluded high-quality evaluations in other languages. We included peer-reviewed evaluations only and therefore did not examine evaluations from organizations and agencies that may focus on intervention implementation rather than scientific research. In standardizing measures of rigor, we constructed a scale based on 13 common measures of methodological quality. Although we accounted for a broad definition of rigor, these measures are based primarily on RCTs as the gold standard. This may overlook the need for flexibility in evaluating structural interventions. Lastly, we limited our definition of "structural interventions" to those that address economic empowerment, education, and substance abuse for HIV prevention. We did so to limit the scope of the review and respond to a mandate from the study funders.

Conclusions

Evaluation of interventions is vital to strengthen the link between good science and sound implementation and to ensure efficient use of limited resources. Rigorous evaluations of structural interventions should include the same components as those of other public health interventions. These components include: identification of outcome(s) of interest, justification of sampling and sample sizes, and utilization of proper analysis techniques [54, 55]. However, evaluations of structural interventions for HIV prevention present additional challenges, including those around evaluation design and implementation, and reporting of results.

Clear development, analysis, and reporting of a TOC is key to understanding the pathways through which structural interventions operate. By identifying, measuring, and reporting intermediary and interacting variables in relation to HIV outcomes, evaluators can better understand why interventions are successful or not. This is particularly useful in instances where the effects of a structural intervention may take years to observe and the cost of long-term studies may restrict measurement of impact. Evaluators may be able to evaluate effects of a structural intervention more holistically and accurately if the variables within a causal pathway are clearly identified. Natural experiments, qualitative methods, and approaches adapted from fields outside epidemiology, such as complexity and systems science, may be more suitable than RCTs for examining and evaluating the complex processes of structural interventions.

Funding This study was funded by the United States Agency for International Development (USAID) under the terms of MEASURE Evaluation cooperative agreement AID-OAA-L-1400004.

Compliance with Ethical Standards

Conflict of interest All authors declares that they have no conflict of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Blankenship KM, Friedman SR, Dworkin S, Mantell JE. Structural interventions: concepts, challenges and opportunities for research. J Urban Health. 2006;83(1):59–72.
2. Parkhurst, J.O. Structural drivers interventions and approaches for prevention of sexually transmitted HIV in general populations: 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 1, and London: UKaid's STRIVE research consortium; 2013.
3. Burström B, Macassa G, Öberg L, Bernhardt E, Smedman L. Equitable child health interventions: the impact of improved water and sanitation on inequalities in child mortality in Stockholm, 1878 to 1925. Am J Public Health. 2005;95(2):208–16.
4. Dreibelbis R, Winch PJ, Leontsini E, Hulland KR, Ram PK, Unicomb L, Luby SP. The integrated behavioural model for water, sanitation, and hygiene: a systematic review of behavioural models and a framework for designing and evaluating behaviour change interventions in infrastructure-restricted settings. BMC Public Health. 2013;13(1):1015.
5. World Health Organization. Global health sector strategy on HIV/AIDS 2011–2015. Geneva: World Health Organization; 2011.
6. Institute of Medicine. Evaluation of PEPFAR. The National Academies; 2013.
7. Pronyk, P., Lutz, B. Policy and interventionme responses for addressing the structural determinants of HIV. Arlington, VA: USAID's AIDS Support and Technical Assistance Resources, AIDSTAR-One, Task Order, 1; 2013.
8. Pettifor AE, Levandowski BA, MacPhail C, Padian NS, Cohen MS, Rees HV. Keep them in school: the importance of education as a protective factor against HIV infection among young South African women. Int J Epidemiol. 2008;37(6):1266–73.
9. Hargreaves JR, Bonell CP, Boler T, Boccia D, Birdthistle I, Fletcher A, Glynn JR. Systematic review exploring time trends in the association between educational attainment and risk of HIV infection in sub-Saharan Africa. AIDS. 2008;22(3):403–14.
10. Jewkes R, Nduna M, Levin J, Jama N, Dunkle K, Puren A, Duvvury N. Impact of stepping stones on incidence of HIV and HSV-2 and sexual behaviour in rural South Africa: cluster randomised controlled trial. BMJ. 2008;337:a506.

11. Cluver L, Boyes M, Orkin M, Pantelic M, Molwena T, Sherr L. Child-focused state cash transfers and adolescent risk of HIV infection in South Africa: a propensity-score-matched case-control study. Lancet Glob Health. 2013;1(6):e362–70.
12. Handa S, Halpern CT, Pettifor A, Thirumurthy H. The government of Kenya's cash transfer intervention reduces the risk of sexual debut among young people age 15–25. PLoS ONE. 2014;9(1):e85473.1101–9.
13. Pettifor A, MacPhail C, Selin A, Gómez-Olivé FX, Rosenberg M, Wagner RG, Wang J. HPTN 068: a randomized control trial of a conditional cash transfer to reduce HIV infection in young women in South Africa—study design and baseline results. AIDS Behav. 2016;20(9):1863–82.
14. Pronyk PM, Kim JC, Abramsky T, Phetla G, Hargreaves JR, Morison LA, Porter JD. A combined microfinance and training intervention can reduce HIV risk behaviour in young female participants. AIDS. 2008;22(13):1659–65.
15. Rosenberg MS, Seavey BK, Jules R, Kershaw TS. The role of a microfinance intervention on HIV risk behavior among Haitian women. AIDS Behav. 2011;15(5):911–8.
16. Ssewamala FM, Han CK, Neilands TB, Ismayilova L, Sperber E. Effect of economic assets on sexual risk-taking intentions among orphaned adolescents in Uganda. Am J Public Health. 2010;100(3):483–8.
17. Latkin C, Weeks MR, Glasman L, Galletly C, Albarracin D. A dynamic social systems model for considering structural factors in HIV prevention and detection. AIDS Behav. 2010;14(2):222–38.
18. Thomas JC, Curtis S, Smith J. The broader context of implementation science [letter]. JAIDS. 2011;58:e19–21.
19. Bonell C, Hargreaves J, Strange V, Pronyk P, Porter J. Should structural interventions be evaluated using RCTs? the case of HIV prevention. Soc Sci Med. 2006;63(5):1135–42.
20. United Kingdom Medical Research Council. Developing and evaluating complex interventions: new guidance. London: United Kingdom Medical Research Council; 2016.
21. Pronyk P, Schaefer J, Somers MA, Heise L. Evaluating structural interventions in public health: challenges, options and global best-practice. Routledge: Structural Approaches in Public Health; 2012.
22. Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, Tyrer P. Framework for design and evaluation of complex interventions to improve health. BMJ. 2000;321(7262):694.
23. World Bank. World Bank Country and Lending Groups; 2016. https://datahelpdesk.worldbank.org/knowledgebase/articles/906519. Accessed June 2016.
24. Jennings L, Gagliardi L. Influence of mhealth interventions on gender relations in developing countries: a systematic literature review. Int J Equity Health. 2013;12(1):1–10. https://doi.org/10.1186/1475-9276-12-85.
25. Kennedy CE, Fonner VA, O'Reilly KR, Sweat MD. A systematic review of income generation interventions, including microfinance and vocational skills training, for HIV prevention. AIDS Care. 2014;26(6):659–73.
26. Baird SJ, Garfein RS, McIntosh CT, Özler B. Effect of a cash transfer interventionme for schooling on prevalence of HIV and herpes simplex type 2 in Malawi: a cluster randomised trial. Lancet. 2012;379(9823):1320–9.
27. Kohler, H., Thornton, R.L. Conditional cash transfers and HIV/AIDS prevention: unconditionally promising? The World Bank Economic Review; 2011.
28. Magnani RJ, McCann HG, Hotchkiss DR, Florence CS. The effects of monetized food aid on reproductive behavior in rural Honduras. Popul Res Policy Rev. 1998;17(4):305–28.
29. Blanchard AK, Mohan HL, Shahmanesh M, Prakash R, Isac S, Ramesh BM, Blanchard JF. Community mobilization, empowerment and HIV prevention among female sex workers in South India. BMC Public Health. 2013;13:234.
30. Ashburn K, Kerrigan D, Sweat M. Micro-credit, women's groups, control of own money: HIV-related negotiation among partnered dominican women. AIDS Behav. 2008;12(3):396–403.
31. Austrian K, Muthengi E. 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. Child Youth Serv Rev. 2014;47:168–75.
32. Ssewamala FM, Ismayilova L. Integrating children's savings accounts in the care and support of orphaned adolescents in rural Uganda. Soc Serv Rev. 2009;83(3):453.
33. Ssewamala FM, Ismayilova L, McKay M, Sperber E, Bannon W, Alicea S. Gender and the effects of an economic empowerment intervention on attitudes toward sexual risk-taking among AIDS-orphaned adolescent youth in Uganda. J Adolesc Health. 2010;46(4):372–8.
34. Cho H, Hallfors DD, Mbai II, Itindi J, Milimo BW, Halpern CT, Iritani BJ. Keeping adolescent orphans in school to prevent human immunodeficiency virus infection: evidence from a randomized controlled trial in Kenya. J Adolesc Health. 2011;48(5):523–6.
35. Hallfors D, Cho H, Rusakaniko S, Iritani B, Mapfumo J, Halpern C. Supporting adolescent orphan girls to stay in school as HIV risk prevention: evidence from a randomized controlled trial in Zimbabwe. Am J Public Health. 2011;101(6):1082–8.
36. Hallfors DD, Cho H, Mbai I, Milimo B, Itindi J. Process and outcome evaluation of a community intervention for orphan adolescents in western Kenya. J Commun Health. 2012;37(5):1101–9.
37. Euser SM, Souverein D, Gowda PRN, Gowda CS, Grootendorst D, Ramaiah R, Kumar S. Pragati: an empowerment interventionme for female sex workers in Bangalore, India. Glob Health Action. 2012;5:19279.
38. Souverein D, Euser SM, Ramaiah R, Gowda PRN, Gowda CS, Grootendorst DC, Kumar S. Reduction in STIs in an empowerment intervention interventionme for female sex workers in Bangalore, India: the Pragati interventionme. Glob Health Action. 2013;6:22943.
39. Swendeman D, Basu I, Das S, Jana S, Rotheram-Borus MJ. Empowering sex workers in India to reduce vulnerability to HIV and sexually transmitted diseases. Soc Sci Med. 2009;69(8):1157–66.
40. Dunbar MS, Kang Dufour MSK, Lambdin B, Mudekunye-Mahaka I, Nhamo D, Padian NS. The SHAZ! project: results from a pilot randomized trial of a structural intervention to prevent HIV among adolescent women in Zimbabwe. PLoS ONE. 2014;9(11):e113621.
41. Baird S, Chirwa E, McIntosh C, Özler B. The short-term impacts of a schooling conditional cash transfer intervention on the sexual behavior of young women. Health Econ. 2010;19(S1):55–68.
42. De Walque D, Dow WH, Nathan R, Abdul R, Abilahi F, Gong E, Krishnan S. Incentivising safe sex: a randomised trial of conditional cash transfers for HIV and sexually transmitted infection prevention in rural Tanzania. BMJ Open. 2012;2(1):e000747.
43. Des Jarlais DC. Structural interventions to reduce HIV transmission among injecting drug users. AIDS. 2000;14:S41–6.
44. Fullilove RE, Green L, Fullilove MT. The Family to Family intervention: a structural intervention with implications for the prevention of HIV/AIDS and other community epidemics. AIDS. 2000;14:S63–7.
45. O'Leary A, Martins P. Structural factors affecting women's HIV risk: a life-course example. AIDS. 2000;14:S68–72.
46. Rotheram-Borus MJ. Expanding the range of interventions to reduce HIV among adolescents. AIDS. 2000;14:S33–40.
47. Schriver B, Mandal M, Muralidharan A, Nwosu A, Dayal R, Das M, Fehringer J. Gender counts: a systematic review of evaluations of gender-integrated health interventions in low-and

middle-income countries. Glob Public Health. 2016;. https://doi.org/10.1080/17441692.2016.1149596.
48. Sumartojo E, Doll L, Holtgrave D, Gayle H, Merson M. Enriching the mix: incorporating structural factors into HIV prevention. AIDS. 2000;14:S1–2.
49. Institute of Medicine Committee on the Social Behavioral Science Base for HIV/AIDS Prevention Intervention Workshop. Assessing the Social and Behavioral Science Base for HIV/AIDS Prevention and Intervention: Workshop Summary: Background Papers. National Academy Press. 1995.
50. Bamberger M, Rao V, Woolcock M. Using mixed methods in monitoring and evaluation: experiences from international development. World Bank Policy Research Working Paper Series. World Bank; 2010.
51. Hipp JR, Morgan SL, Winship C. Counterfactuals and causal inference: methods and principles for social research. JSTOR. 2008.
52. Diez Roux AV. Complex systems thinking and current impasses in health disparities research. Am J Public Health. 2011;101:1627–34.
53. Bamberger M. Integrating quantitative and qualitative research in development projects. World Bank Publications; 2000.
54. Kmet LM, Lee RC, Cook LS. Standard Quality Assessment Criteria for Evaluating Primary Research Papers From a Variety of Fields. Alberta Heritage Foundation for Medical Research; 2004.
55. Rychetnik L, Frommer M, Hawe P, Shiell A. Criteria for evaluating evidence on public health interventions. J Epidemiol Commun H. 2002;56(2):119–27.