



Quantitative Methods Used to Evaluate Impact of Combination HIV Prevention Intervention: A Methodological Systematic Review

Andrainolo Ravalihasy^{1,2,3} · Pearl Anne Ante-Testard^{4,5} · Lidia Kardas-Sloma³ · Yazdan Yazdanpanah⁶ · Manuela De Allegri⁷ · Valéry Ridde^{1,2}

Accepted: 18 January 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Combination HIV prevention aims to provide the right mix of biomedical, behavioral and structural interventions, and is considered the best approach to curb the HIV pandemic. The impact evaluation of combined HIV prevention intervention (CHPI) provides critical information for decision making. We conducted a systematic review of the literature to map the designs and methods used in these studies. We searched original articles indexed in Web of Science, Scopus and PubMed. Fifty-eight studies assessing the impact of CHPI on HIV transmission were included. Most of the studies took place in Asia or sub-Saharan Africa and were published from 2000 onward. We identified 36 (62.1%) quasi-experimental studies (posttest, pretest–posttest and nonequivalent group designs) and 22 (37.9%) experimental studies (randomized designs). The findings suggest that diverse methods are already rooted in CHPI impact evaluation practices as recommended but should be better reported. CHPI impact evaluation would benefit from more comprehensive approaches.

Keywords Combination HIV prevention intervention · Impact evaluation · Decision-making · Evidence-based

Introduction

In 2021, UNAIDS [1] emphasized that the progress in the fight against the HIV/AIDS pandemic is slowing down and is even in jeopardy owing to the effects of COVID-19

crisis on health systems. This statement implies a call for action and the continuation of efforts to curb the pandemic while recognizing that HIV prevention is still a major public health issue. Despite the progress in biomedical prevention research and tools, it is already acknowledged that biomedical approaches alone are not sufficient to curb the epidemic [2]. Besides, numerous behavioral and structural interventions have shown to be effective in improving intermediate outcomes that potentially block the pathway to HIV transmission, such as inducing changes in sexual behaviors. Although such interventions struggle to show any impact on HIV incidence [3, 4], dealing with structural and behavioral determinants of HIV spread alongside the use of biomedical prevention tools is necessary when designing HIV prevention interventions [5, 6]. Therefore, combination HIV prevention, “a dynamic, rights-based approach to providing the right mix of biomedical, behavioral and structural interventions aiming to have the greatest, sustained effort on reducing new HIV infections” [7], is considered the best approach to curb the HIV pandemic [8]. A combination HIV prevention intervention (CHPI) then relates to any intervention that aims to reduce HIV transmission by using strategies that deal with behavioral and structural health determinants supplemented by biomedical prevention tools [5]. Whereas

✉ Andrainolo Ravalihasy
andrainolo.ravalihasy@ird.fr

¹ Centre Population et Développement (Ceped), Institut de Recherche pour le Développement (IRD) et Université Paris Cité, Inserm ERL 1244, 45 Rue Des Saints-Pères, 75006 Paris, France

² French Collaborative Institute On Migrations, CNRS, Aubervilliers, France

³ UMR 1137, Inserm, Université Paris Cité, IAME, Paris, France

⁴ MESuRS and Unité PACRI, Institut Pasteur, Conservatoire National des Arts et Métiers, Paris, France

⁵ Francis I. Proctor Foundation, University of California, San Francisco, San Francisco, CA, USA

⁶ AP-HP, Hôpital Bichat, Service, 75018 Paris, France

⁷ Heidelberg Institute of Global Health, University Hospital and Faculty of Medicine, University of Heidelberg, Heidelberg, Germany

biomedical and behavioral components are individually focused approaches, structural components are designed to affect environmental conditions outside the individuals' control (economic conditions, policies, programmatic vulnerabilities, social inequalities, discrimination, societal norms) [9, 10]. CHPIs add up many components and mobilize all involved parties to account for specific risks and vulnerabilities. By doing so, they take into consideration the contextual needs and conditions of people and communities. CHPIs are expected to prevent HIV transmission by considering their components' effectiveness and relevant hypotheses on how they interact with one other, a credible program impact pathway and a program theory that is able to deal with pragmatic issues [11]. Hence, the impact evaluation of CHPIs raises methodological challenges owing to their multicomponent and complex nature.

Impact evaluations contribute to HIV-related decision making by generating evidence from CHPIs about effective strategies to prevent HIV infections. Indeed, impact evaluations are primarily expected to quantify the extent to which the intervention to be evaluated achieved the intended outcomes. In that sense, they are supposed to establish a causal relationship between the set of activities undertaken during the considered intervention and the improvements of the beneficiaries' circumstances. Beyond the question of whether an intervention is effective, impact evaluations are expected to provide comprehensive evidence that informs the decisions on the implementation, the scale-up, as well as on the continuation or the interruption of an intervention [11, 12]. Currently, impact evaluations are often based on quantitative methods applied in the framework of a Campbellian validity model where the impacts of intervention are quantified in controlled settings (efficacy) and then in real-world settings (effectiveness) [13, 14]. These methods have proven their relevance, which has rooted their use in evaluation practices and has legitimated a form of hierarchy among methods in terms of evidence [15]. However, they present their own limits [14, 16, 17], *inter alia* concerning CHPIs impact evaluation [11]. The current recommendations for CHPIs impact evaluation acknowledge the relevance of diverse quantitative methods and approaches depending on the context of these interventions [11]. For these reasons, this study is conducted to map and critically review the quantitative methods used to assess the efficacy or the effectiveness of CHPIs on HIV transmission. It will help to address the gap between the recommendations about the use of these methods and what is actually occurring.

Methods

The systematic review was conducted according to the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) statement [18] (Online Appendix 1). The protocol was registered and published with PROSPERO (CRD42020210825) and details are presented in Ravalihasy *et al.* [19].

Search Strategy

We searched original articles in English and French indexed in Web of Science, Scopus and PubMed from inception to August 2022. The joint use of these databases is expected to uncover most of the studies relevant to this review [20]. Database-specific search strategies were developed using key terms associated with (i) HIV transmission (prevalence or incidence), (ii) impact of interventions and (iii) prevention and sexual risk exposure [19]. We added terms related to HIV transmission and target populations in order to improve the search strategy (Online Appendix 2). We conducted the literature search in English and then in French after translating text terms (*i.e.* not index terms). All records were retrieved on August 25, 2022, and imported to a reference manager library (Zotero).

Eligibility

A first screening step was conducted based on the title and abstract. References were deemed relevant if:

- The study focused on the impact evaluation of CHPI.
- The analyses were conducted using data gathered from the intervention beneficiaries during these interventions.
- The data allowed to assess the impact of the intervention including the behavioral/structural components.
- HIV incidence, prevalence or averted infections was an intervention outcome.

A second screening step based on full text was then conducted and allowed to exclude all studies that:

- did not assess an impact on HIV transmission (hereafter, irrelevant outcome).
- referred to an intervention that did not include any behavioral or structural component (hereafter, irrelevant intervention).
- where the data did not allow to assess the impact of the behavioral or structural components (hereafter, irrelevant data).

- are not an intervention impact evaluation (hereafter, irrelevant evaluation study).
- were based on simulated data.
- did not report any quantitative method (statistical methods, mathematical modeling or both) to assess the intervention impact on HIV transmission.

In each step, the studies were screened independently by AR and PAA-T. The disagreements were resolved by LKS, MDA or VR.

Data Extraction

A systematic review management software, Covidence (www.covidence.org), was used for data management and extraction. General information (authors, title, date of publication, location where the studies were carried out, purpose and results of the studies) were extracted from the included studies. Specific information about quantitative methods were extracted using a grid developed for this purpose [19]. When necessary, other documents referenced in the full texts, such as the study protocol or pilot study, were used as complementary supports during the data extraction process. Also, each study was classified as an “efficacy” or “effectiveness” study according to the study objectives as stated in the full texts, or according to the fact that the CHPI was already implemented or scaled-up, or using items from the PRECIS-2 tool [21].

Data Analysis

Evaluation Design

Distinctions are made between experimental and quasi-experimental approaches for evaluation design. An experimental design relates to studies where researchers assign the participants to different intervention conditions according to a randomization scheme. A quasi-experimental design relates to studies where researchers do not control intervention allocation or do not use a randomization scheme for intervention allocation [22]. Quasi-experimental designs include posttest design, pretest–posttest design and its extensions (such as interrupted time-series or regression discontinuity), nonequivalent group design, and any combination of the former [23]. Items from the Mixed Methods Appraisal Tool [24, 25] were adapted to extract information about study design reporting (Online Appendix 3).

Statistical Methods

Information about statistical methods were assessed using items developed from the guidelines for Statistical Analyses and Methods in the Published Literature [26]. Information

about sample size are assessed through two items to verify if the data allowed the detection of the expected gain from the CHPIs (expected effect size) with sufficient precision in estimates. The three remaining items allow to check if the studies reported: (i) how the methods fit to the data structure (statistical validity condition), (ii) a measure of precision (confidence and credible intervals) alongside the impact measure, and (iii) how the analysis accounted for the evaluation design [19]. We verified if the studies reported each item and all items together.

Mathematical Modeling

Information about mathematical modeling were extracted using items developed from the guidelines for Strengthening The Reporting of Empirical Simulation Studies [27]. Two items allow the investigation of information about model outputs and their precision. Two items allow the investigation of information about the models’ assumptions. Two items allow the investigation of information about the data used for modeling. One item allows the investigation of information about models implementation [19]. We verified if the studies reported each item and all items together.

Results

The literature search identified 2335 articles, of which 154 were considered relevant based on their titles and abstracts. A total of 58 articles [28–85] concerning 46 CHPI satisfied the inclusion criteria and were included in our review (Fig. 1).

All the 46 considered CHPIs (58 articles) presented behavioral components, of which eighteen (39.1%) presented structural components. Table 1 reports the characteristics of the included studies. Among 53 studies where HIV transmission was a primary outcome, 26 (49.1%) reported a significant reduction of HIV transmission. Among five studies where HIV transmission was a secondary outcome, two (40.0%) reported a significant reduction of HIV transmission. Thirty-six studies (62.1%) including three cost-effectiveness studies were conducted in real-life settings, of which 21 (58.3%) reported a reduction of HIV transmission as intended. Twenty-two studies (37.9%) were conducted in controlled settings, of which seven (31.8%) reported a reduction of HIV transmission as intended. Most of the included studies targeted female sex workers (39.7%), were published after 2000 (94.8%), and were conducted in Asia (36.2%) or in sub-Saharan Africa (44.8%). Data concerning the study characteristics are provided in Online Appendix 4.

Table 2 presents the evaluation design characteristics of the included studies. Among the included studies, 22 (37.9%) used experimental design: 12 (54.6%) were

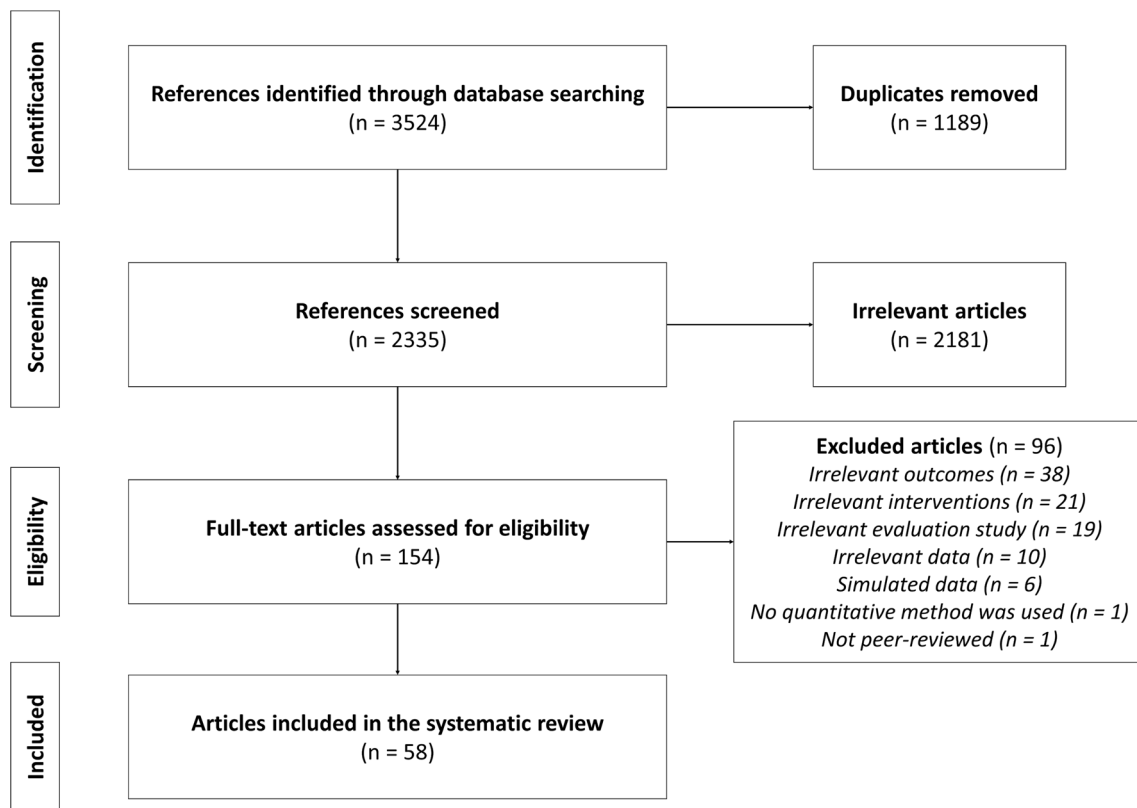


Fig. 1 PRISMA flow diagram for study selection process

cluster-randomized, 9 (40.9%) were individually randomized, and one (4.5%) was a multilevel randomization (i.e., a combination of individual and cluster-randomization). The remaining 36 (62.1%) studies used quasi-experimental designs, including 6 (16.7%) posttest designs, 6 (16.7%) nonequivalent group designs, 8 (22.2%) pretest–posttest designs and 16 (44.4%) combinations of pretest–posttest and nonequivalent group designs. Items concerning sampling (67.2%) and treatment allocation or exposure (82.8%) were more frequently reported than other items. Quasi-experimental studies more frequently reported how confounders and measurement biases were accounted for in the design than in the experimental studies (50.0% vs 36.4%, respectively), but this difference was not significant ($\chi^2 = 1.02$, $p = 0.311$). In particular, among the experimental studies, eight reported blinded procedures: all kinds of outcome assessor [44, 49, 71]; biological outcome assessor [35, 48]; and interviewers [72] or investigators [40, 55] were blinded in these studies. Experimental studies more frequently reported information about intervention administration and adherence than did quasi-experimental studies (63.6% vs 13.9%, respectively, $\chi^2 = 15.3$, $p < 0.001$).

Among the included studies, 51 (87.9%) used statistical methods to assess to what extent the interventions reduced HIV transmission as intended. Among the former,

38 (74.5%) used regression-based methods, twelve (23.5%) used hypothesis testing and one (2.0%) used an analysis of variance. Table 3 presents the statistical methods used by the included studies. The measures of precision and the consideration of evaluation design were the most frequently reported items (80.4%). The latter involved any methods to account for the data generating process [86, 87] (adjustment, stratification, matching, weighting). Confidence and credible intervals were reported in 41 studies (80.4%), while p-values only were reported in five studies (9.8%). Twenty out of twenty-four studies used an expected effect size related to HIV transmission to compute sample size: 15 were experimental and five were quasi-experimental studies. Overall and for each considered item, statistical methods were more frequently reported in experimental studies.

Among the included studies, nine quasi-experimental studies (15.5%) used mathematical models to assess to what extent the interventions reduced HIV transmission as intended. Table 4 presents the reported mathematical models characteristics in the included studies. Information about model implementation were the least reported.

Table 1 Characteristics of the 58 included studies

Study characteristics	n	%
Impact on HIV transmission as primary outcome of the study		
No	5	8.6
Yes	53	91.4
Significant reduction of HIV transmission reported		
No	30	51.7
Yes	28	48.3
Study type (in real life settings)		
Efficacy (in controlled settings)	22	37.9
Effectiveness (in real-world settings)	36	62.1
Study population		
Female sex workers	23	39.7
Drug users	6	10.4
*Female sex workers and drug users	1	1.7
Men who have sex with men	5	8.6
Adolescent and young adults	6	10.4
Businesses employees	1	1.7
Couples	3	5.2
General population	13	22.4
Year of publication		
<2000	3	5.2
2000–2009	19	32.7
2010 onward	36	62.1
Study location		
Asia	21	36.2
Caribbean	1	1.7
Eastern Europe	1	1.7
North America	4	6.9
South/Latin America	3	5.2
Sub-Saharan Africa	26	44.8
**Multi-country	2	3.5

*The intervention was directed toward female sex workers and drug users

**One study was conducted in South Africa; Tanzania; Zimbabwe and Thailand; one study was conducted in USA and in Thailand

Discussion

This study contributes to the literature on CHPI impact evaluation by giving a broad view of the quantitative methods typically being used and how they are reported. Diverse quantitative designs and methods are currently being implemented, depicting the intrinsic complexity and the contexts of these interventions. To derive an intervention impact estimate on HIV transmission, one can use common procedures, opt for specific procedures [88–91] as can be seen in four included studies [36, 38, 44, 50], or develop procedures when relevant as can be seen in one CHPI [92]. Moreover, CHPIs potentially include structural level activities, raising the demand for more comprehensive impact evaluation studies [11]. In our review, two studies [39, 66] used causal pathway analysis (i.e. accounted for the hypothesized relation between intervention components) to derive estimates of CHPI impact. Some other studies adjusted their estimates for intervention implementation outcomes such as coverage and acceptance, or reported such information [58, 67, 75, 93]. In light of the above, diverse methods are already rooted in CHPI impact evaluation practices as recommended [11]. Still, some effort should be made to better report the methods related to the results of these studies in order to best inform theories and practices.

This review identified more quasi-experimental studies than in other reviews on behavioral and structural interventions to prevent HIV infection [94–99]. It is consistent with the fact that such design may be more appropriate to assess the impact on HIV incidence or prevalence [11, 100, 101]. Furthermore, some ethical, political and resources issues as well as the nature of CHPIs make the implementation of randomized designs less feasible [102]. When feasible, some features adaptation pertaining to randomization or blinding procedures often apply, as reported in our review. Indeed, the complex nature of CHPIs challenges the translation into practice of the theoretical properties of randomized designs,

Table 2 Proportion of reported design characteristics in the 58 included studies

Items [19, 25]	Quasi-experimental (N=36)		Experimental (N=22)		Overall (N=58)		χ^2 †	p-value††
	n	%	N	%	n	%		
Sampling	20	55.6	19	86.4	39	67.2	5.88	0.015
Treatment allocation	29	80.6	19	86.4	48	82.8	0.32	0.570
Outcome integrity	5	13.9	3	13.6	8	13.8	0.001	0.978
Consideration of potential confounders and measurement biases	18	50.0	8	36.4	26	44.8	1.02	0.311
Administration/adherence	5	13.9	14	63.6	19	32.8	15.34	<0.001
Reported of all items listed above	1	2.9	0	0.0	1	1.7	0.62	0.430

†Chi-square statistic

††p-value associated with the chi-square statistic

Table 3 Proportion of reported items among 51 studies concerned by statistical methods

Items [19, 26]	Quasi-experimental (N=29)		Experimental (N=22)		Overall (N=51)		χ^2 †	p-value†)
	n	%	n	%	n	%		
Expected difference between groups for sample size calculation	8	27.6	16	72.7	24	47.1	10.23	0.001
Estimate level of precision for sample size calculation	9	31.0	16	72.7	25	49.0	8.70	0.003
Statistical validity condition	3	10.3	4	18.2	7	13.7	0.65	0.421
Reported intervention impacts measures of precision	23	79.3	18	81.8	41	80.4	0.05	0.823
Consideration of evaluation design	19	65.5	22	100.0	41	80.4	9.43	0.002
Reported of all items listed above	0	0.0	3	13.6	3	5.9	4.20	0.04

†Chi-square statistic

††p-value associated with the chi-square statistic

Table 4 Proportion of accurately reported items among 9 studies concerned by mathematical models

Items [19, 27]	n	%
Model outputs	8	88.9
Sensitivity analysis	8	88.9
Model components	7	77.8
Model hypothesis	9	100.0
Data sources	9	100.0
Model inputs	7	77.8
Model implementation	2	22.2
Reported of all items listed above	2	22.2

which shall ensure unbiasedness and precision [103, 104]. These results illustrate the recommendation that no single methodology should be applied as a gold standard to evaluate CHPIs [11], particularly since confounders may impact experimental as well as quasi-experimental designs [105, 106].

Our review contributes to the literature and stands out for giving insights about the methods used to assess the effectiveness of these interventions. Numerous methods are used to assess the impact of CHPIs. This diversity allows to account for evaluation contexts, and practically all types of evaluation designs. Our results shows that although data from one-group posttest designs seem irrelevant for impact evaluation [107], mathematical modeling allows a counterfactual analysis of intervention outcomes. This review highlights that the reported sampling strategy or the data used in most of the quasi-experimental studies and in a few experimental studies was not intended for impact evaluation on HIV transmission. In addition, information about how well the statistical methods in use suit the data structure are not frequently reported. Therefore, the reporting of these methods should be improved in order to make clearer the relevance of the sample in relation to the methods used to derive impact estimates. Indeed, this will help to better

understand the significance of the results given the diversity of the designs and the methods that are actually used.

Implications for Impact Evaluation Studies

This systematic review shows that the availability of diverse approaches, methods and designs allows us to challenge the complexity of CHPIs impact evaluation. The use of pathways analysis may overcome the need for more comprehensive approaches to impact evaluation. In order to go further, the outcomes of impact evaluation should incorporate contextual and implementation outcomes [108]. In that sense, different tools [109, 110] or approaches, such as theory-driven outcomes evaluation [111, 112], may help to enhance impact evaluation designs. By doing so, impact evaluation may improve the generalizability or the transferability of the findings. Nevertheless, while many approaches are already rooted in impact evaluation practices and the future direction for improvement is identified, poor study reporting may hamper the credibility of the findings. The lack of information on design and quantitative methods implementation means we cannot firmly rely on the findings. Many reporting guidelines have been developed to enhance the reporting of health research studies [113–115], and while their use might be cumbersome, their uptake is critical [116].

The lack of reporting observed in this review may be related to major methodological issues [117] given the items that are reported in the data extraction grid. Our results question the sufficiency and the completeness of the procedures of impact evaluation in the included studies, especially concerning the extent to which the data in use and the intervention administration are relevant. Thus, this review points to the need for updating some key principles that guide the planning, the processing, and the reporting of impact evaluation studies in order to take advantage of the strengths and weaknesses of the designs and the methods in use. In light of the above, these principles should deal with three non-exhaustive but essential questions.

First of all, these principles should address the question of the primary recipients of the evaluation findings (e.g. beneficiaries, stakeholders, funders). Although impact evaluation studies share the aim of establishing causal relationship between programs and outcomes, they may have different purposes ranging from testing the relevance of a program within a specific setting to influencing political decisions [12]. Indeed, the evaluation strategy and constraints may differ according to whoever is interested and involved in the evaluation process. It should be clear whether the evaluation studies are intended to only apply and be restituted within the initial program context or to have implications beyond. Indeed, the clarification of this point gives concrete indications on the scope of the evaluation outcomes.

Second, impact evaluation studies should be able to provide information about and account for the data generating process. Here, accounting for data generating process means identifying to what extent the data allows to derive impact estimates and if not, what kind of adjustment are needed. The data generating process constrains certain methodological aspects of the impact evaluation by shaping the data and the sample characteristics, the intervention allocation or exposure, and the confounding and/or the intervention contextual factors. For example, the intervention allocation may constrain the design or the quantitative analysis methods depending on whether the data collection was specifically planned to allow an assessment on the outcome of interest, such as HIV incidence. Moreover, some additional quantitative processes such as power analysis should be performed when the data collection was not planned specifically for impact evaluation purpose.

Third, some implementation outcomes, especially fidelity, should be assessed alongside impacts. Implementation fidelity is a multidimensional concept [108, 118] that encompasses not only the quality of the delivery or the adherence to the intervention, but also the exposure to the intervention, the beneficiaries' responsiveness and the program components differentiation. This outcome deals with theoretical issues such as the program theory and pathways as well as practical issues such as the stakeholders and beneficiaries' participation. Hence, such an outcome constitutes a key intermediate factor for attaining the expected effects from the intervention.

Taking into account these three questions allows to move towards a more comprehensive manner to consider the intervention impact, capturing the efficacy-effectiveness continuum [21]. These questions also allow to introduce the notion of transferability which focuses on a more practical way to consider the generalizability of the findings [110, 119] without questioning the necessity of the probability statements on which the Campbellian generalizability relates to. Indeed, the former one deals with the impact variation that depends on the beneficiaries, the context and the implementation,

while the latter one warrants the relevance of the impact estimates. By addressing these three questions, the evaluation process takes advantage of the intervention theories of action and change [120] and will enable the production of more actionable results.

Strengths and Limitations

To our knowledge, this review is the first to focus on the impact evaluation methodology of CHPI. Hence, our review included different intervention designs and settings, thus limiting the possibility of a sharper methodological analysis of impact evaluation. However, while the grid we developed [19] only include information that are common to each design, it allows to adopt a conservative approach to the analysis by focusing on the items that are reported correctly.

Some references may have been missed by the search equation: we identified few studies published before 2000 and found no study conducted in Western Europe. Nevertheless, we expanded the initial literature search strategy in order to include index terms and text terms that are expected to increase the likelihood of detecting eligible studies. Furthermore the focus on CHPI and HIV transmission, the diversity of included study and the fact that almost half of the included studies did not show a significant impact on HIV transmission is comforting with respect to publication bias [121].

This review also included efficacy as well as effectiveness studies that may have different purposes. Nevertheless, decision makers often equally use evidence from these studies [11, 14]. Also, the specific context of CHPIs may affect the generalizability of the findings in the same way [122].

Modeling studies may account for complementary approaches to impact evaluation [12]. However, these studies were included thanks to their common use in measuring HIV incidence or prevalence and to examine potential intervention impact [123]. Therefore, our study highlights the relevance of mathematical models as tools for CHPI impact evaluation.

Conclusion

This study highlights that diverse methods are already rooted in CHPI impact evaluation practices. Still, some effort have yet to be made to accurately report these methods to allow a better understanding of the findings' significance. In addition, CHPI impact evaluation may benefit from more comprehensive approaches such as path analysis or theory-driven evaluation. Such approaches allow the quantification of the impact of these interventions, while also taking into account the pragmatic issues and causal theories underlying these

interventions. Indeed, the success of a CHPI is supposed to rely on the interaction between the intervention components implementation, and so should the impact evaluation. These findings contribute to inform future directions for impact evaluation practices in order to make available more transferable and generalizable insights into CHPI.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10461-023-04000-8>.

Acknowledgements We thank Elisabeth Adjadj (DESP, Inserm, France) for her assistance in developing the search strategy. We also thank Emilie Brunet and Laurence Goury (IST, IRD, France) for their assistance in finding full texts of articles. We thank Kevin Jean (MESuRS, Cnam, France) for his precious advice and help during the study selection process.

Author Contributions AR designed the study and VR, LKS, YY and MDA contributed to the study design. AR, PAA-T, MDA and LKS performed the literature search. AR and PAA-T extracted the data and conducted the analyses. AR first drafted the manuscript. PAA-T, MDA, VR, LKS and YY critically revised the manuscript. All authors have read and approved the manuscript.

Funding This work was done in the course of the employment of all authors.

Data Availability All data relevant to the study are included in the article or uploaded as supplementary information.

Code Availability Not applicable.

Declarations

Conflict of interest We declare no conflicts of interest.

Ethical Approval Not applicable.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

References

1. Joint United Nations Programme on HIV/AIDS. Unequal, unprepared, under threat: why bold action against inequalities is needed to end AIDS, stop COVID-19 and prepare for future pandemics. Geneva: Joint United Nations Programme on HIV/AIDS; 2021.
2. Padian NS, Buvé A, Balkus J, Serwadda D, Cates W. Biomedical interventions to prevent HIV infection: evidence, challenges, and way forward. *Lancet*. 2008;372:585–99.
3. Gliberman J, Mitra S, Gogolishvili D, Rueda S, Schoffel L, Gangbar K, Shi Q, Rourke SB. HIV/STI prevention interventions: a systematic review and meta-analysis. *Open Med*. 2017;12:450–67.
4. Iskarpatyoti BS, Lebov J, Hart L, Thomas J, Mandal M. Evaluations of structural interventions for HIV prevention: a review of approaches and methods. *AIDS Behav*. 2018;22:1253–64.
5. Kurth AE, Celum C, Baeten JM, Vermund SH, Wasserheit JN. Combination HIV prevention: significance, challenges, and opportunities. *Curr HIV/AIDS Rep*. 2011;8:62–72.
6. Piot P, Bartos M, Larson H, Zewdie D, Mane P. Coming to terms with complexity: a call to action for HIV prevention. *Lancet*. 2008;372:845–59.
7. Joint United Nations Programme on HIV/AIDS. Combination HIV prevention: tailoring and coordinating biomedical. Geneva: Behavioural and Structural Strategies to Reduce New HIV Infections; 2007.
8. Joint United Nations Programme on HIV/AIDS. Fast-tracking combination prevention. Geneva: UNAIDS; 2015.
9. Sumartojo E. Structural factors in HIV prevention: concepts, examples, and implications for research. *AIDS*. 2000;14:S3.
10. Auerbach JD, Parkhurst JO, Cáceres CF. Addressing social drivers of HIV/AIDS for the long-term response: conceptual and methodological considerations. *Glob Public Health*. 2011;6:S293–309.
11. Joint United Nations Programme on HIV/AIDS. Strategic guidance for evaluating HIV prevention programmes. Geneva: UNAIDS; 2010.
12. Gertler PJ, Martinez S, Premand P, Rawlings LB, Vermeersch CMJ. *Why evaluate? Impact Eval. Pract. Second Ed.*; 2016
13. Campbell DT, Stanley JC. *Experimental and quasi-experimental designs for research*. Belmont: Wadsworth; 2011.
14. Chen HT. The bottom-up approach to integrative validity: a new perspective for program evaluation. *Eval Program Plann*. 2010;33:205–14.
15. Cartwright N, Hardie J. *Evidence-ranking schemes, advice guides, and choosing effective policies*. Oxford: Oxford University Press; 2012. p. 135–43.
16. Cartwright N, Hardie J. *What are RCTs good for?* Oxford: Oxford University Press; 2012. p. 122–34.
17. Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. *Soc Sci Med*. 2018;210:2–21.
18. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Syst Rev*. 2021;10:89.
19. Ravalihasy A, Kardaś-Słoma L, Yazdanpanah Y, Ridde V. Quantitative methods used to evaluate impact of health promotion interventions to prevent HIV infections: a methodological systematic review protocol. *Syst Rev*. 2022;11:87.
20. Bramer WM, Rethlefsen ML, Kleijnen J, Franco OH. Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. *Syst Rev*. 2017;6:245.
21. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ*. 2015;350: h2147.
22. de Vocht F, Katikireddi SV, McQuire C, Tilling K, Hickman M, Craig P. Conceptualising natural and quasi experiments in public health. *BMC Med Res Methodol*. 2021;21:32.
23. Reichardt CS. *Quasi-experimentation: a guide to design and analysis*. London: The Guilford Press; 2019.
24. Hong QN, Fàbregues S, Bartlett G, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *Educ Inf*. 2018;34:285–91.
25. Hong QN, Pluye P, Fàbregues S, et al. *Mixed Methods Appraisal Tool (MMAT) version 2018, User guide*; 2018
26. Lang TA, Altman DG. Basic statistical reporting for articles published in Biomedical Journals: The “Statistical Analyses and Methods in the Published Literature” or the SAMPL Guidelines. *Int J Nurs Stud*. 2015;52:5–9.
27. Monks T, Currie CSM, Onggo BS, Robinson S, Kunc M, Taylor SJE. Strengthening the reporting of empirical simulation studies: introducing the STRESS guidelines. *J Simul*. 2019;13:55–67.

28. Alary M, Banandur P, Rajaram SP, Thamattoor UK, Mainkar MK, Paranjape R, Adhikary R, Duchesne T, Isac S, Moses S. Increased HIV prevention program coverage and decline in HIV prevalence among female sex workers in South India. *Sex Transm Dis*. 2014;41:380–7.
29. Baeten JM, Heffron R, Kidoguchi L, et al. Integrated delivery of antiretroviral treatment and pre-exposure prophylaxis to HIV-1-serodiscordant couples: a prospective implementation study in Kenya and Uganda. *PLoS Med*. 2016;13: e1002099.
30. Banandur P, Mahajan U, Potty RS, Isac S, Duchesne T, Abdous B, Ramesh BM, Moses S, Alary M. Population-level impact of Avahan in Karnataka State, South India using multilevel statistical modelling techniques. *JAIDS-J Acquir Immune Defic Syndr*. 2013;62:239–45.
31. Bhave G, Lindan CP, Hudes ES, Desai S, Wagle U, Tripathi SP, Mandel JS. Impact of an intervention on HIV, sexually transmitted diseases, and condom use among sex workers in Bombay, India. *AIDS*. 1995;9(Suppl 1):S21–30.
32. Birdthistle I, Kwaro D, Shahmanesh M, et al. Evaluating the impact of DREAMS on HIV incidence among adolescent girls and young women: a population-based cohort study in Kenya and South Africa. *PLoS Med*. 2021;18: e1003837.
33. Boily M-C, Pickles M, Lowndes CM, et al. Positive impact of a large-scale HIV prevention programme among female sex workers and clients in South India. *AIDS*. 2013;27:1449–60.
34. Chabata ST, Hensen B, Chiyaka T, et al. The impact of the DREAMS partnership on HIV incidence among young women who sell sex in two Zimbabwean cities: results of a non-randomised study. *BMJ Glob Health*. 2021;6: e003892.
35. Coates TJ, Kulich M, Celentano DD, et al. Effect of community-based voluntary counselling and testing on HIV incidence and social and behavioural outcomes (NIMH Project Accept; HPTN 043): a cluster-randomised trial. *Lancet Glob Health*. 2014;2:E267–77.
36. Corbett EL, Makamure B, Cheung YB, Dauya E, Matambo R, Bandason T, Munyati SS, Mason PR, Butterworth AE, Hayes RJ. HIV incidence during a cluster-randomized trial of two strategies providing voluntary counselling and testing at the workplace, Zimbabwe. *AIDS*. 2007;21:483–9.
37. Cowan FM, Pascoe SJ, Langhaug LF, et al. The Regai Dzive Shiri project: results of a randomized trial of an HIV prevention intervention for youth. *AIDS*. 2010;24:2541–52.
38. Doyle AM, Ross DA, Maganja K, et al. Long-term biological and behavioural impact of an adolescent sexual health intervention in Tanzania: follow-up survey of the community-based MEMA kwa Vijana Trial. *PLOS Med*. 2010. <https://doi.org/10.1371/journal.pmed.1000287>.
39. Eaton LA, Kalichman SC, Kenny DA, Harel O. A reanalysis of a behavioral intervention to prevent incident HIV infections: including indirect effects in modeling outcomes of Project EXPLORE. *AIDS Care*. 2013;25:805–11.
40. El-Bassel N, Gilbert L, Terlikbayeva A, et al. Effects of a couple-based intervention to reduce risks for HIV, HCV, and STIs among drug-involved heterosexual couples in Kazakhstan: a randomized controlled trial. *JAIDS-J Acquir Immune Defic Syndr*. 2014;67:196–203.
41. El-Bassel N, McCrimmon T, Mergenova G, et al. A cluster-randomized controlled trial of a combination HIV risk reduction and microfinance intervention for female sex workers who use drugs in Kazakhstan. *J Int AIDS Soc*. 2021;24: e25682.
42. Geidelberg L, Mitchell KM, Alary M, et al. Mathematical model impact analysis of a real-life pre-exposure prophylaxis and treatment-as-prevention study among female sex workers in Cotonou, Benin. *JAIDS J Acquir Immune Defic Syndr*. 2021;86: e28.
43. Ghys PD, Diallo MO, Ettiègne-Traoré V, et al. Effect of interventions to control sexually transmitted disease on the incidence of HIV infection in female sex workers. *AIDS*. 2001;15:1421–31.
44. Go VF, Frangakis C, Minh NL, et al. Efficacy of a multi-level intervention to reduce injecting and sexual risk behaviors among HIV-infected people who inject drugs in Vietnam: a four-arm randomized controlled trial. *PLoS ONE*. 2015. <https://doi.org/10.1371/journal.pone.0125909>.
45. Goswami P, Medhi GK, Armstrong G, et al. An assessment of an HIV prevention intervention among People Who Inject Drugs in the states of Manipur and Nagaland, India. *Int J Drug Policy*. 2014;25:853–64.
46. Grabowski MK, Serwadda DM, Gray RH, et al. HIV prevention efforts and incidence of HIV in Uganda. *N Engl J Med*. 2017;377:2154–66.
47. Gregson S, Adamson S, Papaya S, Mundondo J, Nyamukapa CA, Mason PR, Garnett GP, Chandiwana SK, Foster G, Anderson RM. Impact and process evaluation of integrated community and clinic-based HIV-1 control: a cluster-randomised trial in eastern Zimbabwe. *PLoS Med*. 2007;4:545–55.
48. Hayes RJ, Donnell D, Floyd S, et al. Effect of universal testing and treatment on HIV incidence—HPTN 071 (PopART). *N Engl J Med*. 2019;381:207–18.
49. Hoffman IF, Latkin CA, Kukhareva PV, et al. A peer-educator network HIV prevention intervention among injection drug users: results of a randomized controlled trial in St. Petersburg, Russia. *AIDS Behav*. 2013;17:2510–20.
50. 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.
51. Kagaayi J, Chang LW, Ssempiija V, et al. Impact of combination HIV interventions on HIV incidence in hyperendemic fishing communities in Uganda: a prospective cohort study. *Lancet HIV*. 2019;6:e680–7.
52. Kerrigan D, Mbwambo J, Likindikoki S, et al. Project Shikamana: community empowerment-based combination HIV prevention significantly impacts HIV incidence and care continuum outcomes among female sex workers in Iringa, Tanzania. *JAIDS-J Acquir Immune Defic Syndr*. 2019;82:141–8.
53. Koblin BA. Effects of a behavioural intervention to reduce acquisition of HIV infection among men who have sex with men: the EXPLORE randomised controlled study. *Lancet*. 2004;364:41–50.
54. Konate I, Traore I, Ouedraogo A, et al. Linking HIV prevention and care for community interventions among high-risk women in Burkina Faso-The ARNS 1222 “Yerelon” cohort. *JAIDS-J Acquir Immune Defic Syndr*. 2011;57:S50–4.
55. Latkin CA, Donnell D, Metzger D, et al. The efficacy of a network intervention to reduce HIV risk behaviors among drug users and risk partners in Chiang Mai, Thailand and Philadelphia, USA. *Soc Sci Med*. 2009;68:740–8.
56. Lee MB, Leibowitz A, Rotheram-Borus MJ. Cost-effectiveness of a behavioral intervention for seropositive youth. *AIDS Educ Prev*. 2005;17:105–18.
57. Luchters S, Chersich MF, Rinyiru A, Barasa M-S, King’ola N, Mandaliya K, Bosire W, Wambuga S, Mwarogo P, Temmerman M. Impact of five years of peer-mediated interventions on sexual behavior and sexually transmitted infections among female sex workers in Mombasa, Kenya. *BMC Public Health*. 2008;1:11. <https://doi.org/10.1186/1471-2458-8-143>.
58. Matovu J, Gray R, Makumbi F, Wawer M, Serwadda D, Kigozi G, Sewankambo N, Nalugoda F. Voluntary HIV counseling and testing acceptance, sexual risk behavior and HIV incidence in Rakai, Uganda. *AIDS*. 2005;19:503–11.

59. Matovu JKB, Gray RH, Kiwanuka N, Kigozi G, Wabwire-Mangen F, Nalugoda F, Serwadda D, Sewankambo NK, Wawer MJ. Repeat voluntary HIV counseling and testing (VCT), sexual risk behavior and HIV incidence in Rakai, Uganda. *AIDS Behav.* 2007;11:71–8.
60. Moses S, Plummer FA, Ngugi EN, Nagelkerke NJ, Anzala AO, Ndinya-Achola JO. Controlling HIV in Africa: effectiveness and cost of an intervention in a high-frequency STD transmitter core group. *AIDS.* 1991;5:407–11.
61. Moses S, Ramesh BA, Nagelkerke NJD, et al. Impact of an intensive HIV prevention programme for female sex workers on HIV prevalence among antenatal clinic attenders in Karnataka state, south India: an ecological analysis. *AIDS.* 2008;22:S101–8.
62. Ng M, Gakidou E, Levin-Rector A, Khera A, Murray CJL, Dandona L. Assessment of population-level effect of Avahan, an HIV-prevention initiative in India. *Lancet.* 2011;378:1643–52.
63. Nkambule R, Philip NM, Reid G, et al. HIV incidence, viremia, and the national response in Eswatini: two sequential population-based surveys. *PLoS ONE.* 2021;16: e0260892.
64. Patterson TL, Mausbach B, Lozada R, et al. Efficacy of a brief behavioral intervention to promote condom use among female sex workers in Tijuana and Ciudad Juarez, Mexico. *Am J Public Health.* 2008;98:2051–7.
65. Peak A, Rana S, Maharjan SH, Jolley D, Crofts N. Declining risk for HIV among injecting drug users in Kathmandu, Nepal: the impact of a harm-reduction programme. *AIDS.* 1995;9:1067–70.
66. Pickles M, Boily M-C, Vickerman P, et al. Assessment of the population-level effectiveness of the Avahan HIV-prevention programme in South India: a preplanned, causal-pathway-based modelling analysis. *Lancet Glob Health.* 2013;1:E289–99.
67. Ramesh BM, Beattie TSH, Shajy I, Washington R, Jagannathan L, Reza-Paul S, Blanchard JF, Moses S. Changes in risk behaviours and prevalence of sexually transmitted infections following HIV preventive interventions among female sex workers in five districts in Karnataka state, south India. *Sex Transm Infect.* 2010;86:117–24.
68. Reza-Paul S, Beattie T, Syed HUR, et al. Declines in risk behaviour and sexually transmitted infection prevalence following a community-led HIV preventive intervention among female sex workers in Mysore, India. *AIDS.* 2008;22:S91–100.
69. Roland ME, Neilands TB, Krone MR, Coates TJ, Franses K, Chesney MA, Kahn JS, Martin JN. A randomized noninferiority trial of standard versus enhanced risk reduction and adherence counseling for individuals receiving post-exposure prophylaxis following sexual exposures to HIV. *Clin Infect Dis.* 2011;53:76–83.
70. Sabido M, Giardina F, Hernandez G, Hugo Fernandez V, Ernesto Monzon J, Ortiz R, Montoliu A, Casabona J. The UALE Project: decline in the incidence of HIV and sexually transmitted infections and increase in the use of condoms among sex workers in Guatemala. *JAIDS-J Acquir Immune Defic Syndr.* 2009;51:S35–41.
71. Safren SA, Thomas B, Biello KB, et al. Strengthening resilience to reduce HIV risk in Indian MSM: a multicity, randomised, clinical efficacy trial. *Lancet Glob Health.* 2021;9:e446–55.
72. Sherman SG, Sutcliffe C, Srirojn B, Latkin CA, Aramratanna A, Celentano DD. Evaluation of a peer network intervention trial among young methamphetamine users in Chiang Mai, Thailand. *Soc Sci Med.* 2009;68:69–79.
73. Subramanian T, Ramakrishnan L, Aridoss S, et al. Increasing condom use and declining STI prevalence in high-risk MSM and TGs: evaluation of a large-scale prevention program in Tamil Nadu, India. *BMC Public Health.* 2013. <https://doi.org/10.1186/1471-2458-13-857>.
74. Sweat M, Kerrigan D, Moreno L, Rosario S, Gomez B, Jerez H, Weiss E, Barrington C. Cost-effectiveness of environmental-structural communication interventions for HIV prevention in the female sex industry in the Dominican Republic. *J Health Commun.* 2006;11:123–42.
75. Thilakavathi S, Boopathi K, Girish Kumar CP, et al. Assessment of the scale, coverage and outcomes of the Avahan HIV prevention program for female sex workers in Tamil Nadu, India: is there evidence of an effect? *BMC Public Health.* 2011;11(Suppl 6):S3.
76. Thirumurthy H, Bair EF, Ochwal P, Marcus N, Putt M, Maman S, Napierala S, Agot K. The effect of providing women sustained access to HIV self-tests on male partner testing, couples testing, and HIV incidence in Kenya: a cluster-randomised trial. *Lancet HIV.* 2021;8:e736–46.
77. Thuong NV, Van Nghia K, Hau TP, et al. Impact of a community sexually transmitted infection/HIV intervention project on female sex workers in five border provinces of Vietnam. *Sex Transm Infect.* 2007;83:376–82.
78. Tinajeros F, Miller WM, Castro L, Artiles N, Flores F, Evans JL, Mendoza S, Urquia M, Rodriguez X, Paz-Bailey G. Declining sexually transmitted infections among female sex workers: the results of an HIV and sexually transmitted infection prevention strategy in Honduras, 2006–08. *Int J STD AIDS.* 2012;23:88–93.
79. Traore IT, Meda N, Hema NM, et al. HIV prevention and care services for female sex workers: efficacy of a targeted community-based intervention in Burkina Faso. *J Int AIDS Soc.* 2015. <https://doi.org/10.7448/IAS.18.1.20088>.
80. Vickerman P, Terris-Prestholt F, Delany S, Kumaranayake L, Rees H, Watts C. Are targeted HIV prevention activities cost-effective in high prevalence settings? Results from a sexually transmitted infection treatment project for sex workers in Johannesburg, South Africa. *Sex Transm Dis.* 2006;33:S122–32.
81. Wagman JA, Gray RH, Campbell JC, et al. Effectiveness of an integrated intimate partner violence and HIV prevention intervention in Rakai, Uganda: analysis of an intervention in an existing cluster randomised cohort. *Lancet Glob Health.* 2015;3:e23–33.
82. Wechsberg WM, Zule WA, El-Bassel N, Doherty IA, Minnis AM, Novak SD, Myers B, Carney T. The male factor: Outcomes from a cluster randomized field experiment with a couples-based HIV prevention intervention in a South African township. *Drug Alcohol Depend.* 2016;161:307–15.
83. Williams JR, Alary M, Lowndes CM, et al. Positive impact of increases in condom use among female sex workers and clients in a medium HIV prevalence epidemic: modelling results from Project SIDA1/2/3 in Cotonou, Benin. *PLoS ONE.* 2014. <https://doi.org/10.1371/journal.pone.0102643>.
84. Ye S, Xiao Y, Jin C, Cassell H, Blevins M, Sun J, Vermund SH, Qian H-Z. Effectiveness of integrated HIV prevention interventions among Chinese men who have sex with men: evaluation of a 16-city public health program. *PLoS ONE.* 2012;7: e50873.
85. Yu J, Zhang Y, Jiang J, et al. Implementation of a “county-Township-Village” Allied HIV Prevention and Control Intervention in Rural China. *AIDS Patient Care STDs.* 2017;31:384–93.
86. Pearl J. *Causality: models, reasoning, and inference.* New York: Cambridge University Press; 2000.
87. Pearl J. *Causal inference in statistics: an overview.* *Stat Surv.* 2009;3:96–146.
88. Bennett S, Parpia T, Hayes R, Cousens S. Methods for the analysis of incidence rates in cluster randomized trials. *Int J Epidemiol.* 2002;31:839–46.
89. Hayes RJ, Moulton LH. *Cluster randomised trials, second edition.* 2017. <https://doi.org/10.4324/9781315370286>
90. Murray DM. *Design and analysis of group-randomized trials.* Oxford: Oxford University Press; 1998.
91. Wu Z, Frangakis CE, Louis TA, Scharfstein DO. Estimation of treatment effects in matched-pair cluster randomized trials by

- calibrating covariate imbalance between clusters. *Biometrics*. 2014;70:1014–22.
92. Chandrasekaran P, Dallabetta G, Loo V, et al. Evaluation design for large-scale HIV prevention programmes: the case of Avahan, the India AIDS initiative. *AIDS*. 2008;22:S1.
 93. Ye S, Xiao Y, Jin C, Cassell H, Blevins M, Sun J, Vermund SH, Qian H-Z. Effectiveness of integrated HIV prevention interventions among Chinese men who have sex with men: evaluation of a 16-city public health program. *PLoS ONE*. 2012. <https://doi.org/10.1371/journal.pone.0050873>.
 94. Crepaz N, Tungol-Ashmon MV, Higa DH, et al. A systematic review of interventions for reducing HIV risk behaviors among people living with HIV in the United States, 1988–2012. *AIDS*. 2014;28:633–56.
 95. Lyles CM, Kay LS, Crepaz N, Herbst JH, Passin WF, Kim AS, Rama SM, Thadiparthi S, DeLuca JB, Mullins MM. Best-evidence interventions: findings from a systematic review of HIV behavioral interventions for US populations at high risk, 2000–2004. *Am J Public Health*. 2007;97:133–43.
 96. Ota E, Wariki WM, Mori R, Hori N, Shibuya K. Behavioral interventions to reduce the transmission of HIV infection among sex workers and their clients in high-income countries. *Cochrane Database Syst Rev*. 2011. <https://doi.org/10.1002/14651858.CD006045.pub3>.
 97. Sipe TA, Barham TL, Johnson WD, Joseph HA, Tungol-Ashmon ML, O’Leary A. Structural interventions in HIV prevention: a taxonomy and descriptive systematic review. *AIDS Behav*. 2017;21:3366–430.
 98. Townsend L, Mathews C, Zembe Y. A systematic review of behavioral interventions to prevent HIV infection and transmission among heterosexual, adult men in low-and middle-income countries. *Prev Sci*. 2013;14:88–105.
 99. Wariki WM, Ota E, Mori R, Koyanagi A, Hori N, Shibuya K. Behavioral interventions to reduce the transmission of HIV infection among sex workers and their clients in low- and middle-income countries. *Cochrane Database Syst Rev*. 2012. <https://doi.org/10.1002/14651858.CD005272.pub3>.
 100. Padian NS, McCoy SI, Balkus JE, Wasserheit JN. Weighing the gold in the gold standard: challenges in HIV prevention research. *AIDS*. 2010;24:621–35.
 101. Brookmeyer R. Measuring the HIV/AIDS epidemic: approaches and challenges. *Epidemiol Rev*. 2010;32:26–37.
 102. Petticrew M, Cummins S, Ferrell C, Findlay A, Higgins C, Hoy C, Kearns A, Sparks L. Natural experiments: an underused tool for public health? *Public Health*. 2005;119:751–7.
 103. Stephenson J, Imrie J. Why do we need randomised controlled trials to assess behavioural interventions? *BMJ*. 1998;316:611–3.
 104. Sullivan GM. Getting off the “gold standard”: randomized controlled trials and education research. *J Grad Med Educ*. 2011;3:285–9.
 105. Reichardt CS. Cause and effect. New York: The Guilford Press; 2019. p. 11–25.
 106. Reichardt CS. Threats to validity. New York: The Guilford Press; 2019. p. 26–44.
 107. Reichardt CS. One-group posttest-only designs. New York: The Guilford Press; 2019. p. 94–8.
 108. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, Griffey R, Hensley M. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. *Adm Policy Ment Health Ment Health Serv Res*. 2011;38:65–76.
 109. Cambon L, Minary L, Ridde V, Alla F. A tool to facilitate transferability of health promotion interventions: ASTAIRE. *Sante Publique (Bucur)*. 2014;26:783–6.
 110. Wang S, Moss JR, Hiller JE. Applicability and transferability of interventions in evidence-based public health. *Health Promot Int*. 2006;21:76–83.
 111. Chen HT. The theory-driven approach to outcome evaluation. 2nd ed. Los Angeles: SAGE Publications; 2015. p. 304–39.
 112. Cambon L, Alla F. Understanding the complexity of population health interventions: assessing intervention system theory (ISyT). *Health Res Policy Syst*. 2021;19:95.
 113. Altman DG, Simera I. A history of the evolution of guidelines for reporting medical research: the long road to the EQUATOR Network. *J R Soc Med*. 2016;109:67–77.
 114. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Med*. 2010;7:e1000217.
 115. Simera I, Moher D, Hirst A, Hoey J, Schulz KF, Altman DG. Transparent and accurate reporting increases reliability, utility, and impact of your research: reporting guidelines and the EQUATOR Network. *BMC Med*. 2010;8:24.
 116. Altman DG, Simera I. Responsible reporting of health research studies: transparent, complete, accurate and timely. *J Antimicrob Chemother*. 2010;65:1–3.
 117. Page MJ, McKenzie JE, Higgins JPT. Tools for assessing risk of reporting biases in studies and syntheses of studies: a systematic review. *BMJ Open*. 2018;8:e019703.
 118. Dane AV, Schneider BH. Program integrity in primary and early secondary prevention: are implementation effects out of control? *Clin Psychol Rev*. 1998;18:23–45.
 119. Victora CG, Habicht J-P, Bryce J. Evidence-based public health: moving beyond randomized trials. *Am J Public Health*. 2004;94:400–5.
 120. Chen HT. Logic models and the action model/ change model schema (program theory). 2nd ed. Los Angeles: SAGE Publications; 2015. p. 58–93.
 121. Jessani NS, Williamson RT, Choonara S, Gautier L, Hoe C, Jafar SK, Khalid AF, Rodríguez Salas I, Turcotte-Tremblay A-M, Rodríguez DC. Evidence attack in public health: Diverse actors’ experiences with translating controversial or misrepresented evidence in health policy and systems research. *Glob Public Health*. 2022;0:1–17.
 122. Craig P, Di Ruggiero E, Frohlich KL, et al (2018) Taking account of context in population health intervention research: guidance for producers, users and funders of research. <https://doi.org/10.3310/CIHR-NIHR-01>
 123. McGill E, Er V, Penney T, et al. Evaluation of public health interventions from a complex systems perspective: a research methods review. *Soc Sci Med*. 2021;272: 113697.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.