

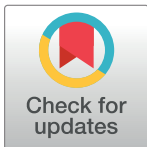
POLICY FORUM

Strengthening implementation guidelines for HIV service delivery: Considerations for future evidence generation and synthesis

Ingrid Eshun-Wilson^{1,2*}, Nathan Ford³, Aaloke Mody¹, Laura Beres⁴, Sheree Schwartz⁵, Stefan Baral⁵, Elvin H. Geng¹

1 Division of Infectious Diseases, School of Medicine, Washington University in Saint Louis, Saint Louis, Missouri, United States of America, **2** Department of Global Health, Stellenbosch University, Cape Town, South Africa, **3** Department of HIV, Viral Hepatitis and Sexually Transmitted Infections, World Health Organization, Geneva, Switzerland, **4** Department of International Health, John Hopkins School of Public Health, Baltimore, Maryland, United States of America, **5** Department of Epidemiology, John Hopkins School of Public Health, Baltimore, Maryland, United States of America

* i.eshun-wilsonova@wustl.edu



OPEN ACCESS

Citation: Eshun-Wilson I, Ford N, Mody A, Beres L, Schwartz S, Baral S, et al. (2023) Strengthening implementation guidelines for HIV service delivery: Considerations for future evidence generation and synthesis. *PLoS Med* 20(3): e1004168. <https://doi.org/10.1371/journal.pmed.1004168>

Published: March 6, 2023

Copyright: © 2023 Eshun-Wilson et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work is supported by the Bill and Melinda Gates Foundation (INV-009840 to EHG), Viiv Healthcare and the National Institute of Health (K24 AI134413 to EHG, and KL2 TR002346 to IEW and AM). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: EHG is an academic editor for Plos Medicines editorial board

Abbreviations: CAG, community adherence group; GRADE, Grading of Recommendations Assessment, Development, and Evaluation; MMS, multi-month scripting; NMA, network meta-analyses; WHO, World Health Organization.

Summary points

- With highly effective diagnostic, prevention, and treatment innovations available in HIV programs globally, the HIV field is increasingly turning to implementation and service delivery questions.
- Developing guidelines for implementation of interventions is markedly challenged by limitations in primary implementation research design and reporting, as well as difficulties in application of evidence synthesis and guideline development tools originally developed to appraise evidence for efficacy.
- Drawing on the processes of developing the WHO HIV service delivery guidelines for testing and treatment between 2018 and 2021, we present challenges and identify areas for future methodological development to improve the incorporation of implementation research across the full spectrum of the evidence generation continuum.
- We highlight gaps in design, measurement, and reporting of primary implementation research, as well as underreporting of relevant program data.
- We describe how routine application of current evidence synthesis tools may not sufficiently answer implementation questions and propose that methodological tools be optimized to identify high-quality non-randomized evidence and reduce penalization for heterogeneity in meta-analysis of implementation research.
- These findings serve as a blueprint for further methodological work to strengthen existing evidence synthesis and guideline development tools for HIV service delivery guidelines and for implementation research more broadly.

Provenance: Not commissioned; externally peer reviewed.

Introduction

Over the last 10 years, global technical guidance for the public health response to HIV—including but not limited to the World Health Organization (WHO)—has sought to make recommendations not only for clinical practice, but also how services should be implemented [1,2]. While the field of HIV has leveraged robust clinical interventions for prevention, diagnosis, and treatment for decades [3,4], much empirical HIV research today is directed at how to alter health systems, optimize performance, and extend the reach of services—that is, implementation research.

The increased focus on implementation questions has highlighted complexities in synthesizing and appraising research for development of scientific guidance on implementation practice [5]. Many of the hallmarks for rigor such as randomization or masking in individual studies or consistency of effects across studies were developed to assess clinical efficacy data. When applied to implementation questions, similar standards may penalize unavoidable procedures in implementation research (e.g., unmasked treatment assignment) as well as fail to assess features of implementation evidence quality such as relevance for implementation in real-world health systems [6]. We need evidence synthesis tools that highlight research domains that assess the setting or population for scientific inferences (considering external validity) as well as threats to validity of those inferences (considering internal validity), as appropriate for implementation research questions [7]. This will ensure that guidance resulting from such syntheses can address “target validity” of the strategies under consideration.

Between 2018 and 2021, the LIVE project (a living database of HIV implementation science research) supported WHO guideline development for HIV through the conduct of systematic reviews and meta-analyses for the 2019 HIV self-testing guidelines and the 2021 HIV treatment service delivery guidelines [8–10]. Review questions and critical outcomes of interest were based on emerging priorities identified during stakeholder-driven processes convened by WHO [11]. Resulting evidence syntheses were presented to guideline development panels comprised of implementing partners from various HIV programs, as well as academics, people living with HIV, and representatives from funder organizations. In this paper, we summarize gaps in primary HIV implementation research methods and reporting identified through this work and propose areas for future methodological development.

Challenges for evidence synthesis and guideline development for HIV implementation research

To supplement recognized evidence synthesis methods during the development of WHO guidelines for HIV service delivery, we applied implementation science tools to systematically characterize research findings. These included the Proctor and colleagues frameworks for characterizing implementation strategies and implementation outcomes and the PRECIS-2 framework for assessing real-world relevance in trials [12–14]. We also re-examined well-recognized systematic review tools, including the Cochrane Risk of Bias tools and the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) framework regarding their application to HIV implementation science questions [15,16]. Through discussion with systematic review teams and WHO guideline developers, a workshop with key HIV implementation scientists, guideline developers and funders, and reviewing relevant literature, we identified several challenges for implementation evidence synthesis and guideline development for HIV service delivery (Table 1).

Table 1. Methodological gaps in evidence synthesis and guideline development for HIV implementation research.

Methodological gap	Detail
Underreporting of implementation strategy characteristics, implementation outcomes, and program data in primary research	<ul style="list-style-type: none"> • Inconsistent or incomplete specification of strategies' components and co-interventions • Infrequent reporting of implementation outcomes (e.g., acceptability, adoption, fidelity) • Mechanisms of action rarely considered or described • Critical contextual factors (that would help generalization) frequently not specified • Informative program data missing from the evidence base
Imperfect fit of evidence synthesis tools for answering implementation questions	<ul style="list-style-type: none"> • Few tools to assess formal applicability to varied real-world (non-trial) settings • Extensive unexplained effect heterogeneity complicates interpretation of pooled estimates • Difficult to establish transitivity in network meta-analysis and obtain indirect estimates
Evidence from implementation research rated "low certainty" using established rating schemas	<ul style="list-style-type: none"> • Unexplained heterogeneity, an expected feature of implementation research, inadvertently reduces evidence certainty ratings • Observational research and natural experiments, potentially rigorous implementation research designs, are penalized in meta-analyses
Limited tools for incorporating implementation research into guideline development frameworks	<ul style="list-style-type: none"> • Limited tools and guidance for evaluation of "indirectness," an evidence appraisal domain highly relevant for establishing external and target validity • Few guidelines for incorporation of quasi-experimental designs and natural experiments into evidence synthesis and guideline development processes • Limited standardized methods for incorporating implementation outcomes (e.g., acceptability, feasibility) into final guideline development decision making

<https://doi.org/10.1371/journal.pmed.1004168.t001>

Suboptimal characterization of strategies, adaptations, mechanisms of action, context, and implementation outcomes in primary HIV implementation research studies

Inconsistent or incomplete specification of strategies' components and co-interventions

Detailed specification and characterization of primary implementation strategies and co-strategies was not robust across studies included in reviews. As an example, in a review of re-engagement in care strategies—actors (those who conducted tracing efforts) and actions (tracing activities performed) were generally well specified; however, other implementation characteristics, such as the dose (frequency and intensity of contact attempts) and temporality (time when tracing was initiated in relation to last visit or appointment) and the action target (how a "lost" patient was defined and how the tracing sought to re-engage them) were less clear [12,17]. In addition, descriptions of adaptations made during implementation were infrequent. A lack of standardized descriptions of implementation characteristics across studies prevented assessment of effect modification by re-engagement strategy components, even qualitatively, and as a result, the ability to develop recommendations for optimal timing or frequency of re-engagement efforts was limited.

Infrequent reporting of implementation outcomes

Studies rarely presented (in primary or ancillary publications) implementation outcomes. Among 8 studies included in a review of community ART initiation, 3 studies reported on cost

of implementation and none reported on other implementation outcomes (i.e., acceptability and appropriateness of implementation strategies, fidelity to delivery method, adoption by health workers and systems, or sustainability) [18]. For guidelines to broadly endorse practice innovations, their effectiveness must be accompanied by such measures—implementation outcomes complement data on effectiveness by contextualizing meta-analytic findings.

Mechanisms of action rarely considered or described

Hypothesized mechanisms by which implementation strategies were expected to exert effects were also rarely characterized. It was rare for protocols, primary or ancillary publications to include any explicit theory of program effectiveness and assumptions—or for studies to share revised mechanistic theories ex-post given new information during study implementation [19–21]. Community adherence groups (CAGs), for example, have been conceptualized as improving HIV treatment outcomes by reducing structural barriers to care through changes in service location and reduced health facility visit frequency, as well as reduce psychosocial barriers through peer support [22–24]. Studies included in a review of reduced visit frequency, however, showed little difference in outcomes for those in CAGs receiving 3-monthly multi-month scripting (MMS) compared to those receiving 3-monthly MMS at a health facility [25,26]—this raises questions about the additional effect of peer support in CAGs (assuming populations and contexts were comparable). Proposing, refining, and synthesizing evidence on mechanisms can help delineate what components are fundamental to the success of a strategy and what components are more contextually specific in their effects.

Critical contextual factors frequently not specified or sufficiently characterized

A further challenge for synthesis was the limited reporting of contextual factors that influence the effectiveness of an implementation strategy—such as the historical and sociopolitical climate, social structures, organizational hierarchies, financing, governance, leadership, human resource capacity, and stigma [27,28]. Understanding how context interacts with intervention exposure and modulates implementation success can facilitate inference in highly varied implementation settings and help identify necessary adaptations for success under new contextual conditions [27,29]. While standardized methods to measure context are still a work in progress conceptually, the eventual development of tools to specify the right kind of context to enable generalization (and knowledge of when generalization is bounded) will aid research utilization and detailed specification of service delivery guidelines [29–31].

Informative program data missing from the evidence base

Implementation evidence generation and synthesis to date has infrequently incorporated extensive program implementation data [32,33]. Routine assessments of publication bias focus on the absence of negative and small trials in the synthesis; for implementation research however, extensive programmatic data is frequently missing from the literature due to programs lacking time, resources, or skill sets to publish reports or scientific manuscripts. Such evidence reflecting real-world implementation is nevertheless important to strengthen the evidence base for future HIV service delivery guidelines.

Imperfect fit of evidence synthesis methods and HIV implementation research

Few tools to assess formal applicability to varied real-world (non-trial) settings

The feasibility of implementing a given intervention within health system constraints is a core component of meaningful guidance for evidence-to-decision processes, yet there are no commonly used tools for assessing whether synthesized study findings are sufficiently pragmatic and relevant to real-world conditions. Thus, understanding to what degree the trial context and conditions mimic real-world conditions has implications for establishing guidance for scale up [34].

To address this gap during guideline review processes, we applied the PRECIS-2 tool to evaluate the extent to which research findings could be applied to routine care [14] (Table 2). The PRECIS-2 tool was originally developed to characterize explanatory or pragmatic trial designs—further refinement and application of such tools for evidence synthesis could inform feasibility determinations for future guideline development processes. The PRECIS-2 framework was not formally used to inform recommendations but provided structure to feasibility guideline development discussions.

Extensive unexplained effect heterogeneity complicates interpretation of pooled effect estimates

Synthesized effect estimates for implementation research frequently do not represent one true underlying effect (or a single distribution of effects). This was demonstrated by high levels of unexplained statistical heterogeneity in our guideline work: the I^2 statistic, measuring statistical heterogeneity was greater than 90% in several meta-analyses, prompting evidence downgrading even in instances where the majority of effect estimates showed benefit (Fig 1). This heterogeneity reflects a well-recognized reality in implementation: strategies inevitably operate differently across contexts. In such instances, where no one true underlying effect measure exists (relevant to all populations and settings), pooled estimates reflect a broader concept of overall benefit or harm [35]. Thus, when considering pooling of estimates from HIV

Table 2. PRECIS-2 tool to assess real-world relevance and feasibility of research [14].

Domain*	Relevant question	Considerations
Eligibility	Who is selected to participate in the trial?	To what extent are the participants in the trial similar to those who would receive this intervention if it was part of usual care?
Recruitment	How are participants recruited into the trial?	How much extra effort is made to recruit participants over and above what that would be used in the usual care setting to engage with patients?
Setting	Where is the trial being done?	How different is the setting of the trial and the usual care setting?
Organization	What expertise and resources are needed to deliver the intervention?	How different are the resources, provider expertise and the organization of care delivery in the intervention arm of the trial and those available in usual care?
Flexibility: delivery	How should the intervention be delivered?	How different are the resources, provider expertise, and the organization of care delivery in the intervention arm of the trial and those available in usual care?
Flexibility: adherence	What measures are in place to make sure participants adhere to the intervention?	How different is the flexibility in how participants must adhere to the intervention and the flexibility likely in usual care?
Follow-up	How closely are participants followed-up?	How different is the intensity of measurement and follow-up of participants in the trial and the likely follow-up in usual care?
Primary outcome	How relevant is it to participants?	To what extent is the trial's primary outcome relevant to participants?
Primary analysis	To what extent are all data included?	To what extent are all data included in the analysis of the primary outcome?

*Domains are graded on a scale of 1 (efficacy focused) to 5 (highly pragmatic).

<https://doi.org/10.1371/journal.pmed.1004168.t002>

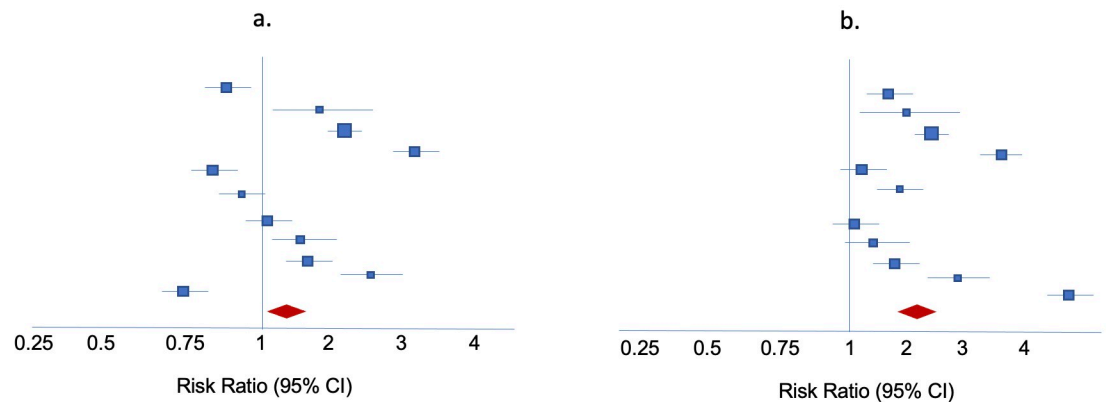


Fig 1. Examples of 2 hypothetical meta-analyses with high statistical heterogeneity but with differing interpretations of overall direction of effect estimates. (a) High statistical heterogeneity with inconsistency in direction of effects (uncertain benefit or harm). (b) High statistical heterogeneity with consistency in direction of effects (benefit). Blue boxes represent effect estimates from individual studies; red diamonds represent pooled effect estimates. Risk ratio greater than 1 represents benefit, less than 1 represents harm [37–40].

<https://doi.org/10.1371/journal.pmed.1004168.g001>

implementation research meta-analyses with high levels of statistical heterogeneity, it is important to assess whether the pooled estimate will be representative or misleading. Future methodological work should formalize additional considerations for implementation research meta-analysis regarding consistency of direction of effects from individual studies, instances where pooling may or may not be valid, as well as ideal interpretation of pooled estimates in light of statistical heterogeneity.

Difficulties establishing transitivity in network meta-analysis to obtain indirect estimates

Unexplained differences in otherwise similarly named strategies, another form of heterogeneity, also complicates the use of network meta-analyses (NMA). NMAs offer a complementary methodology for comparing several implementation strategies through the use of common comparators. In theory, this approach is highly attractive for implementation research where the large number of potential strategies precludes head-to-head comparison of all possibilities. However, the application of this method to implementation research can be challenged by the need to meet the assumption of “transitivity.” In a systematic review of the effects HIVST delivery strategies marked differences in settings and population groups posed a threat to “transitivity” [36]. Transitivity indicates that indirect effects compared through a common group are valid, yet for implementation research, differences in - for example, “standard of care” raises questions about indirect comparisons obtained in such analyses [37]. To address this, networks were separated, and sensitivity and meta-regression analyses are used to explore heterogeneity. It is likely, however, that substantial heterogeneity remains between common comparators [38]. Guidelines and new methods for application of network meta-analysis to implementation research could help to understand effects of interventions where a substantial number of unique implementation strategies have been explored.

Inherent nature of HIV implementation research frequently results in low overall evidence certainty ratings

During the HIV service delivery guideline evidence synthesis process, discussion arose regarding what may be considered good evidence within the GRADE approach and what might be

considered good evidence for implementation. GRADE offers a rigorous and standardized approach for assessing overall evidence certainty for a systematic review outcome (e.g., retention in care) and determining the strength of a guideline recommendation. The application of such rating systems—that value high internal validity (e.g., RCTs) and consistency across effect estimates—to implementation questions, however, means that implementation research can rarely attain a status of “high certainty” evidence, often leading to recommendations that are not strong and, thus, may not be adopted. While this may reflect an overall weak evidence base, another possibility worth interrogating is that the guidance for implementation demands data with different priorities (even if principles are shared).

Observational research and natural experiments, provide critical insights in implementation evidence, but are penalized in meta-analyses

Non-randomized studies—a critical component of HIV implementation research—are initially assigned a low value within GRADE, and unless effect estimates are large or show a dose-response gradient, the inclusion of observational research results in immediate downgrading of evidence certainty [41]. For questions of implementation, however, it is important to consider that findings from observational research may generate estimates that reflect effectiveness during real-world implementation conditions, and their inclusion in evidence synthesis is therefore crucial [42]. The quality of observational research can be highly variable, and further consideration should be given to strengthening tools for distinguishing the methodological quality of non-randomized evidence and adapting guideline development tools to broaden criteria for assigning high-quality evidence ratings.

Unexplained heterogeneity, an inherent feature of implementation research, reduces evidence certainty ratings

Methodological, clinical, and statistical heterogeneity are frequently present in syntheses of implementation research. This frequently results in low evidence quality ratings for implementation research, which limits the ability to make strong implementation guideline recommendations to inform policy. Implementation questions must, however, be tested across heterogeneous settings—the adaptability of the strategy to different settings is what tests its robustness for scalability—and heterogeneity is to be expected [27,43]. Guidelines are needed that outline instances where high-certainty evidence ratings may still be warranted despite heterogeneity.

Limitations of tools and methods for incorporation of implementation research into guideline development frameworks

Limited tools and guidance for evaluation of “indirectness” evidence appraisal domain

Strengthening the guidelines for the GRADE evidence certainty domain of “indirectness” (a measure of the external validity of the synthesized evidence) could improve the application of this domain for implementation questions [7,44]. Guidelines and tools for establishing internal validity are extensive but for external validity, methods are comparatively less well developed, resulting in more subjective assessments. Providing greater detail on requirements within each category of indirectness (i.e., population, intervention, outcome, and indirect comparison) as it relates to questions beyond efficacy can aid review teams to identify if sufficient information is available to establish directness for implementation in the first place and provide more nuanced interpretations. Given the frequent desire for guidelines to translate to diverse and

varied global settings, considerations of expanded descriptions of what data are available from the synthesized evidence to enable inferences in these diverse settings could improve the utility of this domain. Understanding if the mechanism by which a strategy is considered to exert its effects can translate wholly or partially to other settings or populations and could aid in determining where and for whom the evidence is direct or indirect [42]. Incorporation of tools such as the PRECIS-2 tool could strengthen assessments regarding directness or indirectness of evidence [14]. A more in-depth characterization of this GRADE domain for implementation research could help specify conditions under which guideline recommendations are most applicable, as well as inform reporting requirements of primary research to inform future guidelines.

Few guidelines for incorporation of quasi-experimental designs and natural experiments into evidence synthesis and guideline development processes

Quasi-experimental designs are increasingly being used to explore HIV implementation questions. However, they have little recognition in evidence hierarchies and there are few tools to guide incorporation into evidence synthesis [45–49]. For HIV implementation research, quasi-experimental population level effects and programmatic data analyses that can establish causality (if assumptions are met) can contribute real-world program data. Current thinking has argued that natural experiments may be as rigorous as randomized trials because they occur without anyone knowing about the outcomes (they are blinded by nature), and they avoid various artefacts introduced by randomized trials (e.g., so called Hawthorne and John Henry effects). These designs are however limited by the inability to test assumptions for causality and possible bias amplification [50]. Guidelines are emerging for the inclusion of natural experiments in evidence synthesis and clarifying guidance around what constitutes high-quality evidence for natural experiments and how such evidence may be incorporated; this could broaden the evidence base for future HIV service delivery guideline development processes [51].

Limited standardized methods for incorporating implementation outcomes (e.g., acceptability, feasibility) into final guideline development decision-making

There are few detailed methodological guidelines for incorporating the implementation outcomes of acceptability, appropriateness, and adoption into guideline development processes. The incorporation of values and preferences is a cornerstone of the WHO evidence to decision framework; however, specific guidelines for assessments of acceptability, appropriateness, or adoption of strategies for those who are expected to enact or receive the intervention are less clearly defined [52,53]. The Confidence in the Evidence from Reviews of Qualitative research (GRADE-CERQUAL) tool can aid determination of the certainty from qualitative research but with extensive mixed methods and preference research emerging for HIV service delivery, detailed guidance on what evidence, and what questions must be answered to establish “acceptability,” for example, remain unclear [54].

Recommendations for future evidence synthesis for HIV implementation research

This reflection identified several critical areas for improvement in implementation research evidence generation, synthesis, and guideline development (Table 3). Firstly, implementation

Table 3. Recommendations and areas for future methodological development to address evidence generation, synthesis, and guideline development challenges relevant to HIV implementation research.

HIV service delivery evidence continuum [57]	Areas for improvement and future methodological development
Primary implementation research	<ul style="list-style-type: none"> • Report implementation strategy features • Measure and report implementation outcomes • Characterize context and mechanisms of action • Develop simplified platforms for program data reporting
Evidence synthesis	<ul style="list-style-type: none"> • Expand methods for incorporating and assessing observational research and natural experiments • Develop/adapt tools for assessing real-world relevance • Assess/adapt guidelines for pooling heterogeneous effect estimates in implementation research • Expand guidelines for application of network meta-analysis to implementation research
Guideline development	<ul style="list-style-type: none"> • Strengthen methods for incorporating implementation outcomes • Re-assess inconsistency downgrading criteria • Expand tools for establishing indirectness • Adjust evidence rating scores for high-quality non-randomized evidence

<https://doi.org/10.1371/journal.pmed.1004168.t003>

considerations—including characterization of implementation strategies, implementation outcomes, hypothesized mechanisms of action, and context—should be incorporated more broadly in HIV implementation research study methods and dissemination efforts [12,13,55,56]. Second, adapted or new approaches to evidence synthesis that address the limitations as well as the benefits of heterogeneity—in terms of context, implementation strategy, or study design—in addition to the relevance of primary research to real-world settings, are needed [14]. Lastly, guideline-informing evidence rating tools (e.g., GRADE) should determine how rating may differ for questions of implementation as compared to efficacy, particularly in considering the role of high-quality non-randomized research, determinations of indirectness, as well as robust assessment and incorporation of implementation outcomes (e.g., acceptability, adoption, feasibility, sustainability).

The alignment of methods for the production, synthesis, and dissemination is critical for improving the utility and impact of implementation research for the HIV response. International guideline recommendations are routinely based on combinations of expert opinion and research evidence (guided by relevance and quality). If synthesized evidence does not adequately reflect the breadth and depth of implementation efforts, this limits the capacity for guidelines to actually inform translation and real-world implementation efforts. Adapting current evidence generation and synthesis methods to answer efficacy, effectiveness, and implementation questions can support the production of guidelines that provide more comprehensive recommendations on what to implement, how to do it, and under what conditions.

Author Contributions

Conceptualization: Ingrid Eshun-Wilson, Nathan Ford, Aaloke Mody, Elvin H. Geng.

Data curation: Ingrid Eshun-Wilson.

Formal analysis: Ingrid Eshun-Wilson, Elvin H. Geng.

Funding acquisition: Ingrid Eshun-Wilson, Nathan Ford, Elvin H. Geng.

Investigation: Ingrid Eshun-Wilson.

Methodology: Ingrid Eshun-Wilson, Nathan Ford, Elvin H. Geng.

Project administration: Ingrid Eshun-Wilson.

Resources: Ingrid Eshun-Wilson.

Software: Ingrid Eshun-Wilson.

Supervision: Ingrid Eshun-Wilson, Nathan Ford, Elvin H. Geng.

Validation: Ingrid Eshun-Wilson.

Visualization: Ingrid Eshun-Wilson.

Writing – original draft: Ingrid Eshun-Wilson, Nathan Ford, Aaloke Mody, Laura Beres, Sheree Schwartz, Stefan Baral, Elvin H. Geng.

Writing – review & editing: Ingrid Eshun-Wilson, Nathan Ford, Aaloke Mody, Laura Beres, Sheree Schwartz, Stefan Baral, Elvin H. Geng.

References

1. WHO. World Health Organization: Implementation tool for pre-exposure prophylaxis (PrEP). 2022 [cited 2022]. Available from: http://apps.who.int/iris/bitstream/handle/10665/255890/WHO-HIV-2017_19-eng.pdf;jsessionid=ACE3151798ED2A3C720A20476D8DCE69?sequence=1.
2. WHO. World Health Organization: Service Delivery for the treatment and care of people living with HIV. 2021 [cited 2022 Aug 27]. Available from: <https://www.who.int/publications/i/item/9789240023581>.
3. Hargreaves J, Dalal S, Rice B, Anderegg N, Bhattacharjee P, Gafos M, et al. Repositioning Implementation Science in the HIV Response: Looking Ahead From AIDS 2018. *J Acquir Immune Defic Syndr*. 2019; 82.
4. Schmidt HA, Rodolph M, Schaefer R, Baggaley R, Doherty M. Long-acting injectable cabotegravir: implementation science needed to advance this additional HIV prevention choice. *J Int AIDS Soc*. 2022; 25 (7):e25963. <https://doi.org/10.1002/jia2.25963> PMID: 35903882; PubMed Central PMCID: PMC9334859.
5. Hilton Boon M, Thomson H, Shaw B, Akl EA, Lhachimi SK, López-Alcalde J, et al. Challenges in applying the GRADE approach in public health guidelines and systematic reviews: a concept article from the GRADE Public Health Group. *J Clin Epidemiol*. 2021; 135:42–53. Epub 2021/01/22. <https://doi.org/10.1016/j.jclinepi.2021.01.001> PMID: 33476768; PubMed Central PMCID: PMC8352629.
6. Geng EH, Peiris D, Kruk ME. Implementation science: Relevance in the real world without sacrificing rigor. *PLoS Med*. 2017; 14(4):e1002288. <https://doi.org/10.1371/journal.pmed.1002288> PMID: 28441435
7. Westreich D, Edwards JK, Lesko CR, Cole SR, Stuart EA. Target Validity and the Hierarchy of Study Designs. *Am J Epidemiol*. 2019; 188(2):438–443. <https://doi.org/10.1093/aje/kwy228> PMID: 30299451
8. WHO. World Health Organization: Consolidated guidelines on HIV prevention, testing, treatment, service delivery and monitoring: recommendations for a public health approach. 2021 [cited 2022]. Available from: <https://www.who.int/publications/i/item/9789240031593>.
9. WHO. World Health Organization: Guidelines on HIV self-testing and partner notification. 2019 [cited 2022]. Available from: <https://www.who.int/hiv/pub/self-testing/hiv-self-testing-guidelines/en/>.
10. Eshun-Wilson I, Ford N, Le Tourneau N, Baral S, Schwartz S, Kemp C, et al. A Living Database of HIV Implementation Research (LIVE Project): Protocol for Rapid Living Reviews. *JMIR Res Protoc*. 2022; 11(10):e37070. <https://doi.org/10.2196/37070> PMID: 36197704
11. Ford N, Geng E, Ellman T, Orrell C, Ehrenkranz P, Sikazwe I, et al. Emerging priorities for HIV service delivery. *PLoS Med*. 2020; 17(2):e1003028. <https://doi.org/10.1371/journal.pmed.1003028> PMID: 32059023
12. Proctor EK, Powell BJ, McMillen JC. Implementation strategies: recommendations for specifying and reporting. *Implement Sci*. 2013; 8:139. Epub 2013/12/03. <https://doi.org/10.1186/1748-5908-8-139> PMID: 24289295; PubMed Central PMCID: PMC3882890.
13. Proctor EK, Silmere H, Raghavan R, Hovmand P, Aarons GA, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. *Adm Policy Ment Health*. 2011; 38. <https://doi.org/10.1007/s10488-010-0319-7> PMID: 20957426

14. 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. <https://doi.org/10.1136/bmj.h2147> PMID: 25956159
15. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction- GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol*. 2011; 64(4):383–94. Epub 2011/01/05. <https://doi.org/10.1016/j.jclinepi.2010.04.026> PMID: 21195583.
16. Higgins JPT, Thomas J. *Cochrane Handbook for Systematic Reviews of Interventions*. version 6.0 ed. Cochrane; 2019.
17. Mirzazadeh A, Eshun-Wilson I, Thompson RR, Bonyani A, Kahn JG, Baral SD, et al. Interventions to reengage people living with HIV who are lost to follow-up from HIV treatment programs: A systematic review and meta-analysis. *PLoS Med*. 2022; 19(3):e1003940. <https://doi.org/10.1371/journal.pmed.1003940> PMID: 35290369
18. Eshun-Wilson I, Awotiwoon AA, Germann A, Amankwa SA, Ford N, Schwartz S, et al. Effects of community-based antiretroviral therapy initiation models on HIV treatment outcomes: A systematic review and meta-analysis. *PLoS Med*. 2021; 18(5):e1003646. <https://doi.org/10.1371/journal.pmed.1003646> PMID: 34048443
19. Smith JD, Li DH, Rafferty MR. The Implementation Research Logic Model: a method for planning, executing, reporting, and synthesizing implementation projects. *Implement Sci*. 2020; 15(1):84. <https://doi.org/10.1186/s13012-020-01041-8> PMID: 32988389
20. Dixon-Woods M, Bosk CL, Aveling EL, Goeschel CA, Pronovost PJ. Explaining Michigan: developing an ex post theory of a quality improvement program. *Milbank Q*. 2011; 89(2):167–205. <https://doi.org/10.1111/j.1468-0009.2011.00625.x> PMID: 21676020.
21. Geng EH, Baumann AA, Powell BJ. Mechanism mapping to advance research on implementation strategies. *PLoS Med*. 2022; 19(2):e1003918. <https://doi.org/10.1371/journal.pmed.1003918> PMID: 35134069
22. Luque-Fernandez MA, Van Cutsem G, Goemaere E, Hilderbrand K, Schomaker M, Mantangana N, et al. Effectiveness of patient adherence groups as a model of care for stable patients on antiretroviral therapy in Khayelitsha, Cape Town, South Africa. *PLoS ONE*. 2013; 8(2):e56088. Epub 2013/02/19. <https://doi.org/10.1371/journal.pone.0056088> PMID: 23418518; PubMed Central PMCID: PMC3571960.
23. Umberto Pellecchia SB, Spencer Nundwe, Andy Bwanali, Bote Zamadenga, Metcalf CA, Helen Bygrave, et al. “We are part of a family”. Benefits and limitations of community ART groups (CAGs) in Thyolo, Malawi: a qualitative study. *J Int AIDS Soc*. 2017;(20):21374. <https://doi.org/10.7448/IAS.20.1.21374> PMID: 28406273
24. Decroo T, Koole O, Remartinez D, dos Santos N, Dezembro S, Jofrisse M, et al. Four-year retention and risk factors for attrition among members of community ART groups in Tete, Mozambique. *Trop Med Int Health*. 2014; 19(5):514–21. Epub 2014/06/06. <https://doi.org/10.1111/tmi.12278> PMID: 24898272.
25. Fatti G, Ngorima-Mabhena N, Mothibi E, Muzenda T, Choto R, Kasu T, et al. Outcomes of Three- Versus Six-Monthly Dispensing of Antiretroviral Treatment (ART) for Stable HIV Patients in Community ART Refill Groups: A Cluster-Randomized Trial in Zimbabwe. *J Acquir Immune Defic Syndr*. 2020; 84(2):162–72. Epub 2020/02/26. <https://doi.org/10.1097/QAI.0000000000002333> PMID: 32097252; PubMed Central PMCID: PMC7172979.
26. Tukei BB, Fatti G, Tiam A, Ngorima-Mabhena N, Tukei VJ, Tshabalala I, et al. Twelve-Month Outcomes of Community-Based Differentiated Models of Multimonth Dispensing of ART Among Stable HIV-Infected Adults in Lesotho: A Cluster-Randomized Noninferiority Trial. *J Acquir Immune Defic Syndr*. 2020; 85(3):280–91. Epub 2020/07/16. <https://doi.org/10.1097/QAI.0000000000002439> PMID: 32665460.
27. Edwards N, Barker PM. The importance of context in implementation research. *J Acquir Immune Defic Syndr*. 1999; 2014(67 Suppl 2):S157–S162. <https://doi.org/10.1097/qai.0000000000000322> PMID: 25310123.
28. Hontelez JAC, Bulstra CA, Yakusik A, Lamontagne E, Bärnighausen TW, Atun R. Evidence-based policymaking when evidence is incomplete: The case of HIV programme integration. *PLoS Med*. 2021; 18(11):e1003835. <https://doi.org/10.1371/journal.pmed.1003835> PMID: 34752470
29. Mehrotra ML, Petersen ML, Geng EH. Understanding HIV Program Effects: A Structural Approach to Context Using the Transportability Framework. *J Acquir Immune Defic Syndr*. 2019; 82 Suppl 3(Suppl 3):S199–S205. <https://doi.org/10.1097/QAI.0000000000002202> PMID: 31764255.
30. Pfadenhauer LM, Gerhardus A, Mozygemba K, Lysdahl KB, Booth A, Hofmann B, et al. Making sense of complexity in context and implementation: the Context and Implementation of Complex Interventions

- (CICI) framework. *Implement Sci.* 2017; 12(1):21. <https://doi.org/10.1186/s13012-017-0552-5> PMID: 28202031
31. May CR, Johnson M, Finch T. Implementation, context and complexity. *Implement Sci.* 2016; 11(1):141. <https://doi.org/10.1186/s13012-016-0506-3> PMID: 27756414
 32. Rucinski K, Masankha Banda L, Olawore O, Akolo C, Zakaliya A, Chilongozi D, et al. HIV Testing Approaches to Optimize Prevention and Treatment for Key and Priority Populations in Malawi. *Open Forum Infect Dis.* 2022; 9(4):ofac038. Epub 20220307. <https://doi.org/10.1093/ofid/ofac038> PMID: 35265725; PubMed Central PMCID: PMC8900928.
 33. Olawore O, Astatke H, Lillie T, Persaud N, Lyons C, Kamali D, et al. Peer Recruitment Strategies for Female Sex Workers Not Engaged in HIV Prevention and Treatment Services in Côte d'Ivoire: Program Data Analysis. *JMIR Public Health Surveill.* 2020; 6(4):e18000. Epub 20201001. <https://doi.org/10.2196/18000> PMID: 33001039; PubMed Central PMCID: PMC7563635.
 34. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. *Med Care.* 2012; 50. <https://doi.org/10.1097/MLR.0b013e3182408812> PMID: 22310560
 35. Deeks JJ, Higgins J, Altman DG. *Cochrane Handbook for Systematic Reviews of Interventions*. Chapter 10: Analysing data and undertaking meta-analyses. 6.2 ed2021.
 36. Eshun-Wilson I, Jamil MS, Witzel TC, Glidded DV, Johnson C, Le Trouneau N, et al. A Systematic Review and Network Meta-analyses to Assess the Effectiveness of Human Immunodeficiency Virus (HIV) Self-testing Distribution Strategies. *Clin Infect Dis.* 2021; 73(4):e1018–e1028. <https://doi.org/10.1093/cid/ciab029> PMID: 34398952
 37. Chaimani A, Caldwell DM, Li T, Higgins JPT, Salanti G. Chapter 11: Undertaking network meta-analyses. *Cochrane Handbook for Systematic Reviews of Interventions* version 62 2019.
 38. Geng EH, Petersen ML. Network meta-analyses: powerful but not without perils. *Lancet HIV.* 2014; 1(3):e95–6. Epub 2015/10/02. [https://doi.org/10.1016/S2352-3018\(14\)00010-1](https://doi.org/10.1016/S2352-3018(14)00010-1) PMID: 26424122.
 39. Jamil MS, Eshun-Wilson I, Witzel TC, Siegfried N, Figueroa C, Chitembo L, et al. Examining the effects of HIV self-testing compared to standard HIV testing services in the general population: A systematic review and meta-analysis. *EClinicalMedicine* 2021; 38:100991. <https://doi.org/10.1016/j.eclinm.2021.100991> PMID: 34278282
 40. Witzel TC, Eshun-Wilson I, Jamil MS, Tilouche N, Figueroa C, Johnson CC, et al. Comparing the effects of HIV self-testing to standard HIV testing for key populations: a systematic review and meta-analysis. *BMC Med.* 2020; 18(1):381. <https://doi.org/10.1186/s12916-020-01835-z> PMID: 33267890
 41. Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P, et al. GRADE guidelines: 9. Rating up the quality of evidence. *J Clin Epidemiol.* 2011; 64(12):1311–1316. <https://doi.org/10.1016/j.jclinepi.2011.06.004> PMID: 21802902
 42. Rothwell PM. External validity of randomised controlled trials: “to whom do the results of this trial apply?”. *Lancet (London, England).* 2005; 365(9453):82–93. [https://doi.org/10.1016/S0140-6736\(04\)17670-8](https://doi.org/10.1016/S0140-6736(04)17670-8) PMID: 15639683.
 43. Bulstra CA, Hontelez JAC, Otto M, Stepanova A, Lamontagne E, Yakusik A, et al. Integrating HIV services and other health services: A systematic review and meta-analysis. *PLoS Med.* 2021; 18(11): e1003836. Epub 20211109. <https://doi.org/10.1371/journal.pmed.1003836> PMID: 34752477; PubMed Central PMCID: PMC8577772.
 44. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 8. Rating the quality of evidence—indirectness. *J Clin Epidemiol.* 2011; 64(12):1303–1310. <https://doi.org/10.1016/j.jclinepi.2011.04.014> PMID: 21802903
 45. Oladele EA, Badejo OA, Obanubi C, Okechukwu EF, James E, Owihonda G, et al. Bridging the HIV treatment gap in Nigeria: examining community antiretroviral treatment models. *J Int AIDS Soc.* 2018; 21(4):e25108. Epub 2018/04/21. <https://doi.org/10.1002/jia2.25108> PMID: 29675995; PubMed Central PMCID: PMC5909112.
 46. Bor J, Fox MP, Rosen S, Venkataramani A, Tanser F, Pillay D, et al. Treatment eligibility and retention in clinical HIV care: A regression discontinuity study in South Africa. *PLoS Med.* 2017; 14(11): e1002463. <https://doi.org/10.1371/journal.pmed.1002463> PMID: 29182641
 47. Mody A, Sikazwe I, Czaicki NL, Wa Mwanza M, Savory T, Sikombe K, et al. Estimating the real-world effects of expanding antiretroviral treatment eligibility: Evidence from a regression discontinuity analysis in Zambia. *PLoS Med.* 2018; 15(6):e1002574. <https://doi.org/10.1371/journal.pmed.1002574> PMID: 29870531
 48. Mody A, Sikazwe I, Namwase AS, Wa Mwanza M, Savory T, Mwila A, et al. Effects of implementing universal and rapid HIV treatment on initiation of antiretroviral therapy and retention in care in Zambia: a natural experiment using regression discontinuity. *Lancet HIV.* 2021. Epub 2021/10/18. [https://doi.org/10.1016/S2352-3018\(21\)00186-7](https://doi.org/10.1016/S2352-3018(21)00186-7) PMID: 34656208.

49. Reeves BC, Deeks JJ, Higgins JPT, Shea B, Tugwell P, Wells GA. Chapter 24: Including non-randomized studies on intervention effects. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. *Cochrane Handbook for Systematic Reviews of Interventions* version 6.2 (updated February 2021). 2021.
50. Craig P, Katikireddi SV, Leyland A, Popham F. Natural Experiments: An Overview of Methods, Approaches, and Contributions to Public Health Intervention Research. *Annu Rev Public Health*. 2017; 38(1):39–56. <https://doi.org/10.1146/annurev-publhealth-031816-044327> PMID: 28125392
51. Bärnighausen T, Tugwell P, Røttingen JA, Shemilt I, Rockers P, Geldsetzer P, et al. Quasi-experimental study designs series-paper 4: uses and value. *J Clin Epidemiol*. 2017; 89:21–9. Epub 20170330. <https://doi.org/10.1016/j.jclinepi.2017.03.012> PMID: 28365303.
52. Luoto J, Maglione MA, Johnsen B, Chang C. A Comparison of Frameworks Evaluating Evidence for Global Health Interventions. *PLoS Med*. 2013; 10(7):e1001469. <https://doi.org/10.1371/journal.pmed.1001469> PMID: 23874159
53. Moberg J, Oxman AD, Rosenbaum S, Schünemann HJ, Guyatt G, Flottorp S, et al. The GRADE Evidence to Decision (EtD) framework for health system and public health decisions. *Health Res Policy Syst*. 2018; 16(1):45. Epub 2018/05/31. <https://doi.org/10.1186/s12961-018-0320-2> PMID: 29843743; PubMed Central PMCID: PMC5975536.
54. Lewin S, Glenton C, Munthe-Kaas H, Carlsen B, Colvin CJ, Gulmezoglu M, et al. Using qualitative evidence in decision making for health and social interventions: an approach to assess confidence in findings from qualitative evidence syntheses (GRADE-CERQual). *PLoS Med*. 2015; 12(10):e1001895. <https://doi.org/10.1371/journal.pmed.1001895> PMID: 26506244; PubMed Central PMCID: PMC4624425.
55. Nilsen P, Bernhardsson S. Context matters in implementation science: a scoping review of determinant frameworks that describe contextual determinants for implementation outcomes. *BMC Health Serv Res*. 2019; 19(1):189. <https://doi.org/10.1186/s12913-019-4015-3> PMID: 30909897
56. Lewis CC, Klasnja P, Powell BJ, Lyon AR, Tuzzio L, Jones S, et al. From Classification to Causality: Advancing Understanding of Mechanisms of Change in Implementation Science. *Front Public Health*. 2018; 6:136. Epub 2018/06/06. <https://doi.org/10.3389/fpubh.2018.00136> PMID: 29868544; PubMed Central PMCID: PMC5949843.
57. Graham ID, Logan J, Harrison MB, Straus SE, Tetroe J, Caswell W, et al. Lost in knowledge translation: time for a map? *J Contin Educ Heal Prof*. 2006; 26(1):13–24. Epub 2006/03/25. <https://doi.org/10.1002/chp.47> PMID: 16557505.