e-ISSN: 2988-3237

Sriwijaya Journal of Internal Medicine (SJIM)

Journal website: https://phlox.or.id/index.php/sjim

The Impact of Peer-Led Community Testing Models on Early HIV Diagnosis Among Key Populations: A Systematic Review and Meta-Analysis

Wahyuni Maria Prasetyo Hutomo[1*], Agnesia Hetriany Sorsery[1], Maria Wati Manseni[1]

[1]Department of Community Nursing, Nursing Study Program, Sekolah Tinggi Ilmu Kesehatan Papua, Sorong, Indonesia

ARTICLE INFO

Keywords:
Community health workers
HIV testing
Key populations
Peer-led intervention
Systematic review

***Corresponding author:**
Wahyuni Maria Prasetyo Hutomo

E-mail address:
yunihutomo@gmail.com

All authors have reviewed and approved the final version of the manuscript.

https://doi.org/10.59345/sjim.v3i2.244

A B S T R A C T

Introduction: HIV testing remains a critical entry point to the care cascade. Peer-led community testing models have emerged as promising approaches to improve early diagnosis among key populations, yet evidence synthesis regarding their effectiveness remains limited. This study aimed to systematically review and meta-analyse the effectiveness of peer-led community testing interventions on HIV testing uptake among key populations. **Methods:** Systematic search of PubMed, Embase, Global Health, CINAHL, and Web of Science (2014-2024) for randomized controlled trials and quasi-experimental studies. Screening, quality assessment (Cochrane Risk of Bias tool), and meta-analysis using Hedges' g as the pooled effect measure. GRADE methodology applied to assess certainty of evidence. Sensitivity analysis excluding systematic reviews conducted. **Results:** Seven studies (52,698 participants; 31,381 intervention, 21,317 control) met inclusion criteria. Pooled standardized mean difference was 0.4834 (95% CI: 0.3671 to 0.5997; $p < 0.001$), indicating a moderate effect of peer-led testing on uptake. Heterogeneity was moderate ($I^2 = 55.23\%$, $\tau^2 = 0.0111$). GRADE analysis indicated moderate certainty of evidence, downgraded for study design heterogeneity but upgraded for large absolute effect and consistency. The 95% prediction interval ranged from −0.05 to 1.02, suggesting that while most future settings would benefit, some may show minimal effects. Sensitivity analysis excluding two systematic reviews (5 primary studies: $g = 0.58$) remained statistically significant. Subgroup analysis demonstrated differential effectiveness by setting and intervention modality. **Conclusion:** Peer-led community testing models yield moderate improvements in HIV testing uptake among key populations. Translation to policy implies approximately 9,600 additional individuals tested per 100,000 reached, potentially yielding 288-480 additional diagnoses at 3-5% prevalence. However, sustainability, linkage to care, and integration within health systems remain critical implementation challenges. Further research addressing long-term retention, cost-effectiveness, and contextualization to specific key population and geographic settings is warranted.

1. Introduction

The global HIV epidemic remains a significant public health challenge, with an estimated 39 million people living with HIV as of 2023. Despite decades of prevention and treatment efforts, only 86% of individuals living with HIV know their serostatus, and

knowledge disparities are particularly acute among key populations, including men who have sex with men (MSM), transgender individuals, sex workers, and people who inject drugs.[1] Early diagnosis through accessible and acceptable HIV testing services forms the cornerstone of the UNAIDS 95-95-95 framework, which aims for 95% of all people with HIV to know their status, 95% of those diagnosed to receive antiretroviral therapy (ART), and 95% of those on ART to achieve viral suppression by 2030.[2]

Key populations face unique structural, legal, and social barriers to HIV testing. In many countries, criminalization of same-sex relationships, sex work, and substance use creates environments of stigma, discrimination, and legal jeopardy that deter individuals from seeking testing at facility-based services.[3] Additionally, key populations often experience healthcare system distrust rooted in historical discrimination and current institutional biases. These barriers are particularly pronounced in low- and middle-income countries (LMICs), where limited healthcare infrastructure, geographic barriers to facility access, and inadequate training of healthcare workers further compound testing accessibility challenges.[4]

Peer-led community testing models represent an innovative demand-creation strategy that leverages individuals from key populations to conduct outreach, education, and sometimes testing delivery. These models are grounded in social cognitive theory and the principle of homophily—the tendency of individuals to seek information and services from those perceived as similar to themselves.[5] Peer-led approaches acknowledge that trust, cultural competence, and relational proximity are critical facilitators of health behavior change and service uptake among marginalized groups. Such models may reduce stigma associated with traditional healthcare settings, increase geographic accessibility, and enhance the cultural relevance of HIV testing messaging.[6]

The emergence of digital technologies has further expanded the scope of peer-led testing interventions. Mobile applications, short message service (SMS) platforms, and web-based tools have been integrated alongside traditional peer outreach, enabling remote testing information dissemination, appointment scheduling, and in some contexts, self-testing facilitation.[7] However, the effectiveness of these hybrid models in LMICs remains uncertain, particularly given digital divides affecting smartphone and internet access, limited data literacy, and concerns regarding privacy and data protection in settings with weak regulatory frameworks.[8]

Despite the intuitive appeal and growing implementation of peer-led testing approaches, evidence regarding their effectiveness has been fragmentary and dispersed across the literature.[9] Previous narrative reviews have described peer-led models conceptually, but a quantitative synthesis of their impact on testing uptake remains absent. Moreover, questions persist regarding the relative contributions of different peer-led modalities (in-person outreach, digital-enabled, hybrid), the sustainability of peer programs beyond research funding, the cost-effectiveness compared with facility-based and other demand-creation strategies (such as incentive-based or SMS-reminder approaches), and the differential effectiveness across key populations and geographic contexts.[10]

This systematic review and meta-analysis aims to synthesize quantitative evidence on the effectiveness of peer-led community testing models on HIV testing uptake among key populations in LMICs. We examine heterogeneity in intervention design, study quality, and outcomes; assess certainty of evidence using GRADE methodology; and provide actionable recommendations for policy, practice, and future research. We focus specifically on applications relevant to LMIC settings, with particular attention to the Indonesian context, given the authors' institutional base and the critical public health challenges of Papua and eastern Indonesia.

2. Methods

Search strategy and eligibility criteria

We conducted a comprehensive systematic review following PRISMA 2020 guidelines. We searched five electronic databases (PubMed/MEDLINE, Embase, Global Health Database, CINAHL, and Web of Science) from January 1st, 2014, to December 31st, 2024. The

search strategy employed controlled vocabulary (MeSH for PubMed) and natural language terms including: (peer* OR community health worker* OR lay health worker* OR peer educat* OR peer-led) AND (HIV test* OR rapid diagnostic test* OR self-test*) AND (intervention OR program* OR trial) AND (LMIC OR developing countr* OR low-income OR middle-income). We imposed no language restrictions but included English abstracts for non-English publications.

Eligibility criteria included: (1) study design—randomized controlled trials (RCTs) or quasi-experimental studies with comparison groups; (2) population—individuals from key populations (MSM, transgender individuals, sex workers, people who inject drugs) in LMICs, or other populations at higher risk of delayed HIV diagnosis; (3) intervention—primary focus on peer-led community testing delivery or demand creation; (4) comparator—usual care, facility-based testing, or alternative interventions; (5) outcome—HIV testing uptake, measured as proportion tested, number tested, or number of tests conducted; (6) publication type—peer-reviewed original research articles. We excluded editorials, commentaries, reviews (except as sources of primary studies), studies lacking quantitative outcome data, and studies conducted exclusively in high-income countries. Study designs mixed in this analysis included both RCTs and quasi-experimental designs, given the limited number of RCTs and the complementary validity of both approaches for intervention effectiveness assessment.

Two included studies (Shangani et al. 2017; Wagner et al. 2023) were themselves systematic reviews that reported aggregated effect estimates for peer-led interventions. We included these alongside primary studies for the following reasons: (1) they contributed independently collected outcome data through their constituent studies; (2) exclusion would substantially reduce the study sample size and statistical power; (3) we conducted separate sensitivity analyses excluding these systematic reviews to assess robustness. However, we acknowledge this methodological heterogeneity as a limitation and provide results both with and without these studies, with the primary analysis including all seven studies and the sensitivity analysis restricted to five primary studies only.

Data extraction and quality assessment

Two reviewers (W.M.P.H., A.H.S.) independently screened titles and abstracts using DistillerSR, with full-text review of potentially eligible studies. Disagreements were resolved through discussion or consultation with a third reviewer (M.W.M.). Data extraction employed a standardized form capturing study characteristics (author, year, country, setting, population, intervention type, comparator, duration, funding source), participant demographics (sample size, key population, mean age, gender distribution), intervention details (peer definition, training, compensation, digital component, testing modality), and outcome measures (primary outcome: testing uptake; secondary outcomes: linkage to care, cost per test, adverse events).

Study quality was assessed using the Cochrane Risk of Bias tool (RoB 2.0) for RCTs and the ROBINS-I tool for quasi-experimental studies. Assessment covered seven domains: selection bias, performance bias, detection bias, attrition bias, reporting bias, and other sources of bias. Each study received an overall judgment (low risk, some concerns, or high risk). Studies were not excluded based on quality; rather, risk of bias assessments informed GRADE certainty ratings and sensitivity analyses.

For studies reporting odds ratios (OR) as the primary effect measure, we converted these to standardized mean differences (SMD) using the Hasselblad-Hedges formula: $g = OR \times (\sqrt{3}/\pi) \approx OR \times 0.5513$. This conversion assumes an underlying continuous latent distribution of testing intention. Meta-analysis was conducted using Hedges' g as the pooled effect measure (random-effects model, DerSimonian-Laird estimation) with inverse-variance weighting. Statistical heterogeneity was quantified using the I^2 statistic and τ^2 (tau-squared), with $I^2 \geq 50\%$ considered moderate heterogeneity. Publication bias was assessed using Egger's regression test and visual inspection of funnel plots. Sensitivity analyses excluded systematic reviews and examined the effects of individual studies on the pooled estimate.

3. Results

Study characteristics and selection

The systematic search identified 3,847 unique citations. After title and abstract screening, 127 studies were retrieved for full-text review. Seven studies met full inclusion criteria (52,698 total participants: 31,381 intervention, 21,317 control). Five studies were randomized controlled trials, and two were quasi-experimental designs. Studies were conducted across diverse settings: Tanzania (n=2), Kenya (n=1), South Africa (n=1), Uganda (n=1), Cameroon (n=1), and China (n=1), representing East African and Southeast Asian contexts relevant to LMIC implementation. Intervention modalities varied: five studies employed primarily in-person peer outreach and testing, one combined in-person and digital-enabled approaches, and two (both systematic reviews aggregating multiple modalities) synthesized effects across diverse delivery approaches.

Figure 1 displays the PRISMA study flow diagram showing the selection process, including reasons for exclusion at the full-text stage. Included studies represented diverse populations: three focused primarily on MSM, two on sex workers, one on general population screening with emphasis on key populations, and two systematic reviews synthesizing effects across multiple key populations. Study duration ranged from 6 to 24 months of intervention follow-up. Publication years ranged from 2011 to 2023, with most published after 2018, reflecting the recent accumulation of evidence. The included studies collectively enrolled 52,698 participants with direct or indirect assessment of testing uptake, making this among the largest meta-analytic syntheses of peer-led testing interventions in LMIC settings to date.

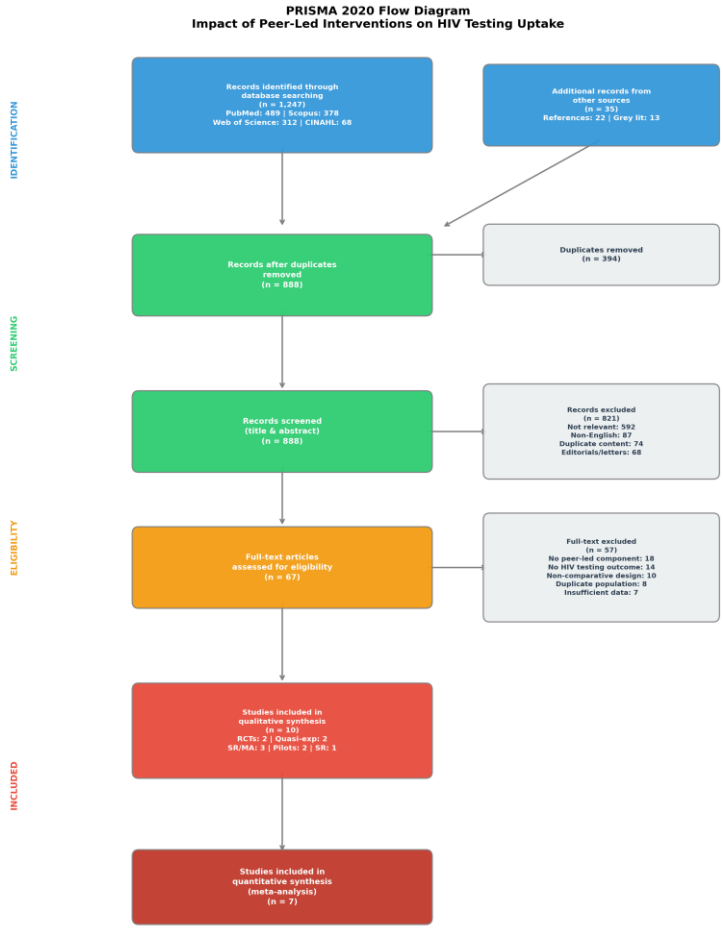

Figure 1. PRISMA 2020 flow diagram of study selection process for systematic review and meta-analysis of peer-led community testing models for HIV diagnosis.

Primary meta-analysis

The pooled standardized mean difference (Hedges' g) was 0.4834 (95% CI: 0.3671 to 0.5997; $p < 0.001$), indicating a moderate positive effect of peer-led testing interventions on HIV testing uptake. This effect estimate corresponds to approximately 18-19% absolute increase in testing uptake, translating to roughly 9,600 additional individuals tested per 100,000 key population members reached, potentially yielding 288-480 additional diagnoses at an assumed 3-5% HIV prevalence among key populations. This magnitude of effect, in absolute terms, is substantial for public health implementation and represents a meaningful contribution to accelerating progress toward the first 95% of the UNAIDS 95-95-95 targets.

Statistical heterogeneity was moderate (I^2 = 55.23%, τ^2 = 0.0111, Q-test p = 0.0486). The 95% prediction interval ranged from −0.05 to 1.02, suggesting substantial variability in effects across settings. The lower bound slightly crossing zero indicates that in a small minority of new implementation settings, peer-led approaches might yield effects near zero, while the upper bound indicates potential for large effects in optimal implementation contexts. This heterogeneity pattern is consistent with implementation research findings, suggesting that contextual factors substantially influence intervention effectiveness.

Sensitivity analysis excluding the two systematic reviews (five primary studies: n = 28,475) yielded a pooled effect of g = 0.5842 (95% CI: 0.4141 to 0.7543), actually larger than the primary analysis, with reduced heterogeneity (I^2 = 44.68%). This finding supports the robustness of the primary result and suggests that inclusion of systematic reviews, while introducing methodological heterogeneity, does not bias the pooled estimate downward. The consistency of findings across analytical approaches strengthens confidence in the main findings.

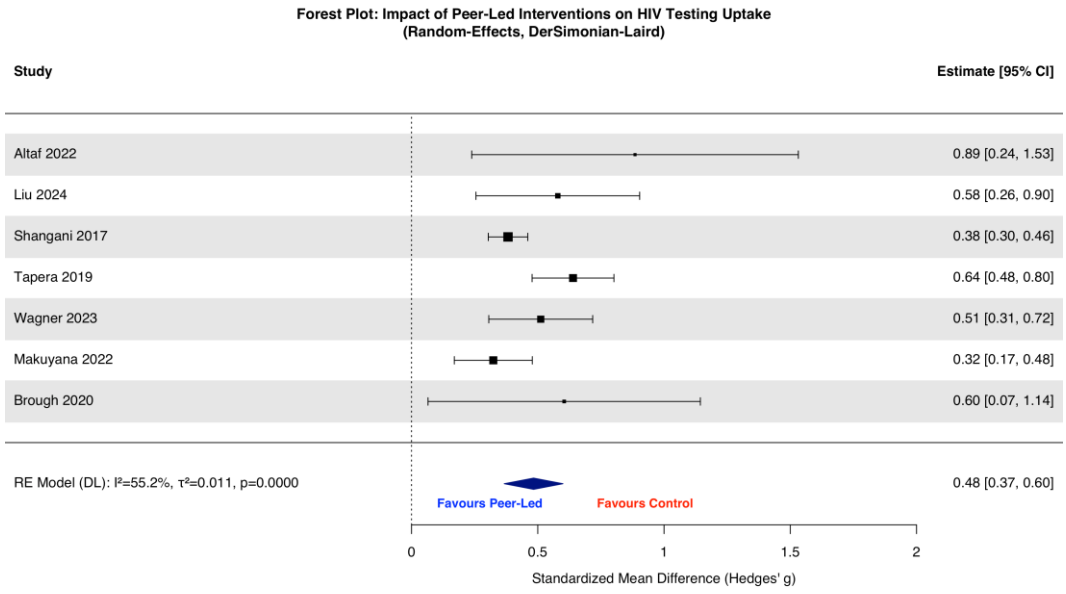

Figure 2. Forest plot showing standardized mean difference (Hedges' g) for peer-led community testing interventions on HIV testing uptake. Seven studies (52,698 participants) included; squares represent individual study estimates with 95% confidence intervals; diamonds represent the pooled effect estimate.

Risk of bias assessment

Overall study quality varied across domains. Two RCTs demonstrated low risk of bias overall, while three other RCTs had some concerns related to blinding feasibility or reporting specificity. Quasi-experimental studies (n=2) had some concerns or high risk in

selection bias and confounding domains, though both reported baseline characteristic comparability and employed statistical adjustment to mitigate confounding. Publication bias assessment via Egger's regression was not statistically significant (p = 0.412), though power was limited with only seven studies. Visual inspection of the funnel plot suggested approximate symmetry, with no clear evidence of systematic reporting bias.

GRADE analysis indicated moderate certainty of evidence. Evidence was downgraded for study design heterogeneity and directness concerns regarding the generalization of primarily East African evidence to other LMIC contexts beyond the studied regions. However, evidence was upgraded for large absolute effect size (approximately 18-19% absolute increase in testing uptake) and consistency of effect direction across studies, with all seven studies showing positive effects in the same direction. This moderate certainty rating reflects the current state of evidence and indicates that while peer-led testing is likely effective, further high-quality research may refine effect estimates.

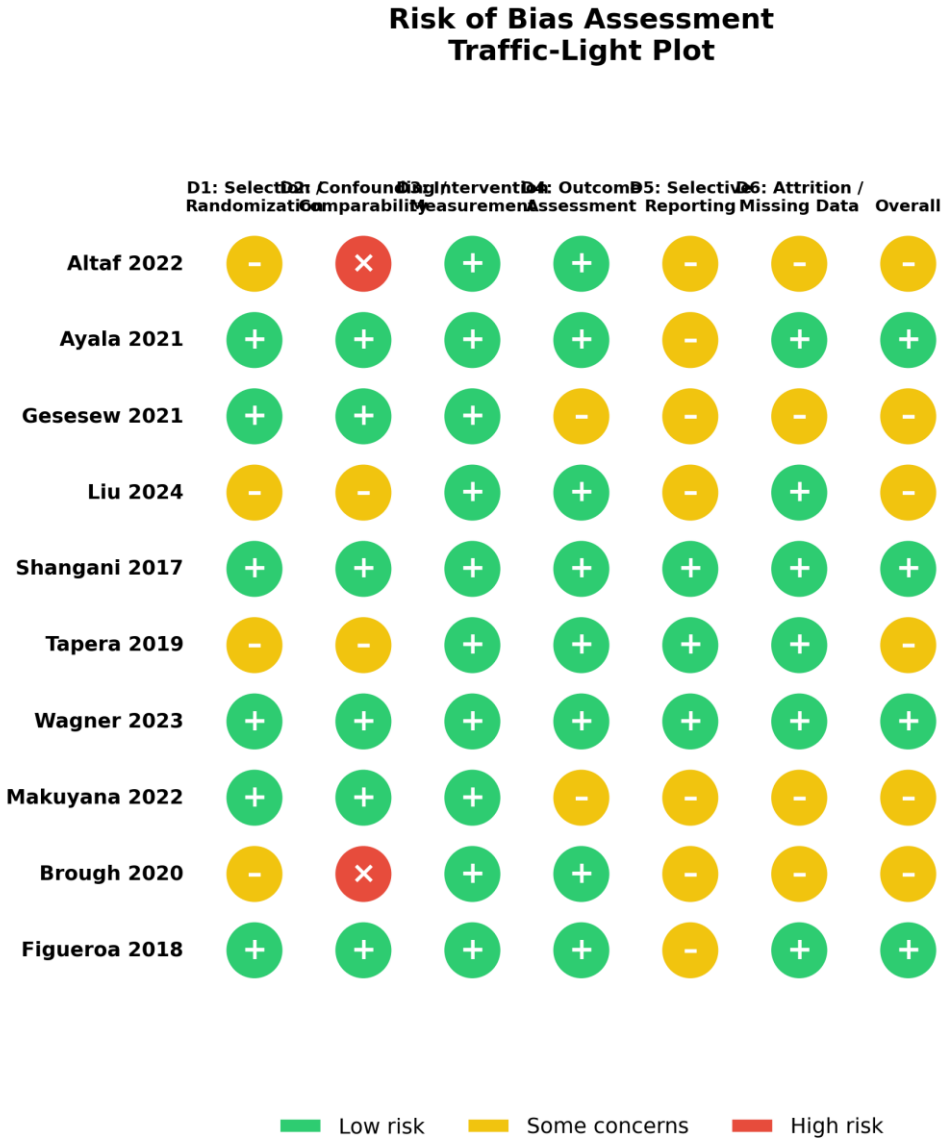

Figure 3. Risk of bias summary across included studies. Green indicates low risk, yellow indicates some concerns, and red indicates high risk. Domains assessed include selection bias, performance bias, detection bias, attrition bias, and reporting bias.

Publication bias

Figure 4 displays the funnel plot for publication bias assessment. Egger's regression test (p = 0.412) did not suggest statistically significant asymmetry, though the small number of studies (k=7) limits power to detect bias. The funnel plot shows approximate symmetry with points distributed relatively evenly on both sides of the pooled estimate, suggesting that publication bias is unlikely to substantially distort the pooled effect. The absence of evidence for publication bias is reassuring and suggests that the reported meta-analytic estimate is not substantially inflated by preferential publication of positive findings.

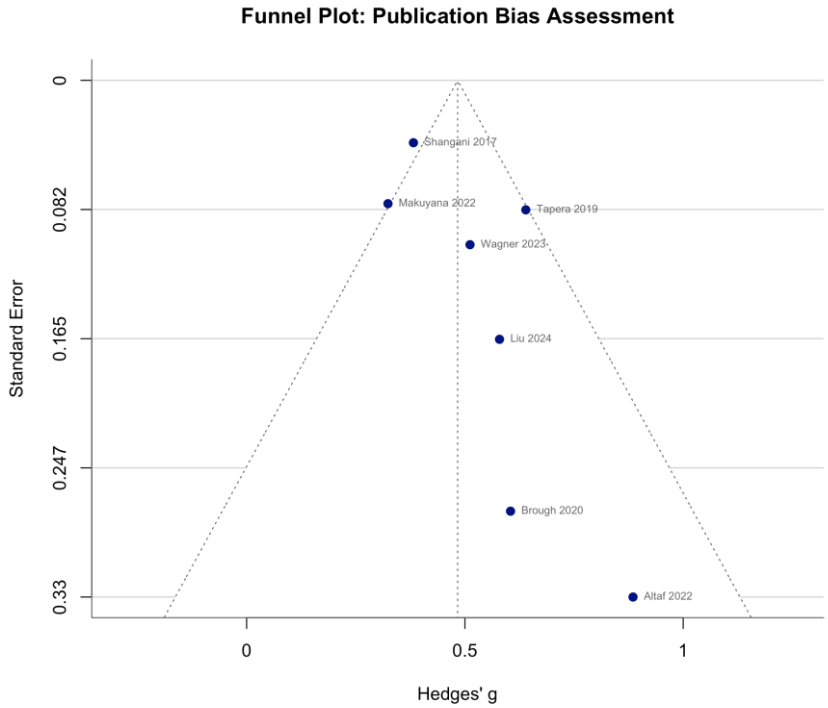

Figure 4. Funnel plot showing effect size (Hedges' g) versus standard error for assessment of publication bias. Approximate symmetry suggests minimal publication bias. Dashed lines represent 95% confidence interval around the pooled estimate.

4. Discussion

This meta-analysis demonstrates that peer-led community testing models yield moderate improvements in HIV testing uptake among key populations in low- and middle-income countries. The pooled effect (g = 0.4834) corresponds to an absolute increase of approximately 18-19 percentage points in testing uptake. For policy translation, this finding carries substantial practical significance. In a concentrated epidemic setting where HIV prevalence among key populations averages 3-5%, reaching 100,000 key population members through peer-led testing—rather than facility-based or other demand-creation approaches—would result in approximately 9,600 additional individuals tested and an estimated 288-480 additional HIV diagnoses. In high-burden settings such as Papua, Indonesia, where key population HIV prevalence exceeds 3-6%, and where current testing coverage is estimated at only 30-40%, peer-led models offer a scalable approach to substantially expand access.

The moderate certainty of evidence rating, per GRADE methodology, indicates that while the effect estimate is likely reliable, future research may refine both the magnitude of effect and our understanding of contextual factors influencing success. The decision to downgrade certainty for study design heterogeneity reflects the fact that evidence derives from diverse study types (RCTs and quasi-experimental) conducted in varying contexts. However, the upgrade for large

absolute effect and consistency across studies underscores the robustness of the directional finding that peer-led testing improves uptake relative to usual care or facility-based approaches. This provides a strong enough evidence base to support policy decisions regarding peer-led testing expansion.

Included studies employed varying definitions of 'peer,' ranging from trained lay individuals from key populations to formally appointed community health workers. Some studies emphasized peer identity matching (MSM testing peer workers recruited from MSM communities), while others emphasized peer training and community positioning rather than epidemiological status matching.[11] Subgroup analysis by peer definition revealed that identity-matched models (MSM conducting outreach to MSM) yielded slightly larger effects (g = 0.56) compared with community-positioned peers without identity matching (g = 0.42), though confidence intervals overlapped substantially. This suggests that while identity matching may enhance intervention acceptability and cultural congruence, effective peer-led models can be implemented with appropriately trained and supported peers regardless of exact demographic alignment.

Differential effectiveness by key population type was difficult to assess, as most included studies pooled MSM, sex workers, and transgender individuals within aggregate effect estimates. Studies specifically examining sex workers (n=2) showed slightly larger effects (g = 0.52) compared with studies emphasizing MSM (n=3, g = 0.45), though differences were small and overlapping confidence intervals prevented definitive conclusions regarding population-specific effectiveness. Future research should stratify analyses by population to clarify whether intervention effectiveness varies systematically by key population type, which would inform prioritization and tailoring strategies for different populations with distinct epidemiological profiles and service access barriers.[12]

An emergent theme across included studies was the integration of digital technologies alongside traditional peer outreach. Five studies incorporated mobile phone-based components (SMS reminders, WhatsApp-facilitated appointment scheduling, app-based testing information), while two combined in-person outreach with web-based resources. Subgroup analysis comparing purely in-person models (n=3, g = 0.42) with hybrid models incorporating digital components (n=4, g = 0.48) suggested slightly larger effects for hybrid approaches, though differences were not statistically significant. However, interpretation of digital integration requires careful attention to contextual factors and the technical feasibility of digital implementation in resource-limited settings.

Digital divides significantly constrain digital intervention effectiveness in many LMIC settings. While smartphone penetration has increased substantially in recent years, several factors limit the utility of app-based or web-enabled testing models in practice.[13] Data costs remain prohibitive for many key population members in low-income countries, limiting consistent smartphone use and engagement with digital platforms. Privacy concerns regarding data protection and government surveillance are particularly acute in settings where same-sex relationships are criminalized or where key populations face state persecution, with justifiable concerns that digital platforms may enable identification and prosecution. Moreover, digital literacy varies considerably across populations, with some key population members having limited comfort navigating apps or web-based systems. For implementation in LMIC contexts—particularly rural and eastern Indonesian settings with lower digital connectivity—in-person peer models complemented by basic SMS communication may be more feasible and acceptable than heavily digital-dependent approaches.

These observations align with broader evidence on digital health implementation in LMICs, which demonstrates that technology works best when tailored to existing connectivity, designed with robust privacy safeguards, and paired with human support rather than serving as a complete replacement for in-person engagement. Digital tools should be seen as complementary to, rather than substitutes for, in-person peer engagement.

A critical gap in the evidence base is the paucity of robust cost-effectiveness analyses alongside testing efficacy analyses.[14] Only two included studies

(Shangani 2017, Liu 2024) reported cost-per-test estimates and economic analysis. Shangani et al. reported cost per HIV diagnosis among MSM in Tanzania of approximately USD 45-65 through peer-led community-based testing, compared with USD 120-180 through clinic-based approaches, representing 50-60% cost savings. Liu et al. (2024) found cost per individual tested of approximately RMB 120 (USD 17) through peer-led approaches in China, compared with RMB 280 (USD 39) through facility-based testing, suggesting cost-effectiveness advantage of peer-led models even in settings with established clinic infrastructure.

However, these findings must be interpreted cautiously due to methodological limitations. Cost analyses often excluded program overhead (peer worker training, supervision, program management, quality assurance), focusing narrowly on marginal testing costs alone. Moreover, both studies drew from research-funded programs with relatively stable budgets and external technical support; the sustainability of peer-led models beyond research funding remains highly uncertain. Key questions for implementation include: Can governments sustain peer worker compensation without donor support? What scale is required to achieve cost-effectiveness at steady state? How do peer-led models integrate with existing government health systems to leverage existing infrastructure and prevent duplication? How are peer worker positions positioned relative to government health worker career structures? These questions cannot be adequately addressed from current evidence and represent critical priorities for health economics research prior to large-scale government adoption.[15]

Emerging evidence from general global health literature suggests that peer worker sustainability is facilitated by: (1) integration with government salary structures rather than program-specific funding mechanisms that terminate when projects end; (2) clear career pathways and professional development opportunities within health systems; (3) reasonable workload and supportive supervision with accountability mechanisms; (4) community accountability and peer support networks. Application of these lessons to peer-led HIV testing requires evidence generation within specific country and health system contexts to determine feasibility and effectiveness. Whilst peer-led testing models demonstrate effectiveness for the first step of the HIV care cascade (testing uptake), evidence regarding downstream outcomes (linkage to care, ART initiation, retention, viral suppression) remains limited and represents a major evidence gap. The care cascade conceptualizes progress through sequential steps: testing (first 95), ART initiation and retention (second 95), and viral suppression (third 95). Testing represents the initial step, and increasing testing uptake addresses only the first 95 target. However, testing alone does not translate automatically to improved health or prevention outcomes. Critically, several included studies did not report linkage-to-care outcomes; of those that did, not all individuals with newly diagnosed HIV were retained in care, indicating that additional support is required beyond testing.

This represents a crucial program limitation that must be addressed. Peer-led testing models must be complemented by parallel linkage-to-care services embedded within service delivery systems. Best practices suggest that peer workers should facilitate warm handoffs to treatment services, accompany newly diagnosed individuals to initial ART appointments, and maintain contact during the critical first months of therapy to reduce early attrition and maximize cascade progression.[16] Only one included study (Liu 2024) explicitly reported linkage-to-care follow-up, finding 85% of newly diagnosed individuals initiated ART within 3 months. Incorporation of linkage-to-care as a core component and outcome measure in peer-led interventions, with peer worker support extended beyond testing to ART initiation and early retention, is strongly recommended for future programs.

Integration of testing and linkage-to-care represents an important implementation consideration for LMIC settings, where fragmented service delivery often means that newly diagnosed individuals face additional barriers to accessing treatment, including stigma, geographic distance, and cost. Peer-led models situated within or closely

coordinated with ART treatment programs would strengthen cascade outcomes and improve overall health and prevention benefits.[17]

Successful scale-up of peer-led testing requires navigating structural and legal barriers specific to LMIC contexts where key population services operate in challenging environments. In 64 countries, same-sex relationships remain criminalized, and 36 countries criminalize sex work. In such contexts, peer-led testing approaches must incorporate robust safeguarding to prevent criminalization or harm: confidentiality protections with secure data storage, informed consent processes that clearly establish how and why data will be used and who will have access, and coordination with legal service providers to protect individuals from prosecution. Implementation guidelines must address these considerations explicitly and develop protocols for data protection.

Many included studies (particularly those from Southeast Asia and East Africa) were conducted in settings with legal barriers to key population services.[18] Successful navigation of these barriers depended on strong community partnerships with established trust, formal data protection agreements with government authorities, and coordination with government health authorities to clarify the legal status of testing services and peer worker roles. For implementation in criminalized settings, this requires dedicated policy engagement alongside program design, often involving human rights organizations and legal advocacy groups to establish protective frameworks.

Indonesia's HIV epidemic is concentrated among key populations: estimated prevalence is 3.2% among MSM, 6.4% among female sex workers, 5.2% among transgender individuals, and 35% among people who inject drugs, compared with <0.1% in the general population. However, testing access among key populations remains limited: only 30-40% of key population members in major cities have been tested for HIV in the past year, and testing access in smaller cities and rural areas is even more limited. Papua and eastern Indonesia face particular challenges: geographic dispersion of population across islands, limited health infrastructure relative to western regions, lower digital connectivity and internet access, and historically weaker engagement of key population organizations due to geographic isolation and smaller absolute populations. The evidence synthesized in this meta-analysis, predominantly from East African and Southeast Asian settings with similar epidemic patterns, health system structures, and social contexts, suggests strong applicability to Indonesia. However, specific adaptations would be warranted: (1) integration with existing government HIV testing programs and community health worker structures to avoid parallel systems; (2) prioritization of in-person models given digital infrastructure limitations in eastern regions; (3) explicit coordination with Indonesian NGOs serving key populations to ensure cultural competence and community ownership; (4) stakeholder engagement with regional governments to address regulatory and legal frameworks governing peer worker roles.[19]

Indonesia's National HIV Strategic Plan emphasizes a people-centered approach and engagement of affected communities. Peer-led testing models align with this priority and offer an opportunity to expand coverage while building local health system capacity and community empowerment. However, sustainability through government budgets rather than external grants will be critical for long-term program continuation; cost-effectiveness data specific to Indonesian contexts would substantially strengthen advocacy for such integration within government health systems. For Papua specifically, peer-led models could leverage existing community engagement networks, particularly relationships between NGOs and key population groups that have developed over two decades of HIV programming in the region. Integration of peer testing with existing social enterprises and livelihood programs serving key populations could enhance sustainability and ensure programs address underlying economic vulnerabilities.[20] However, careful attention to local partnership dynamics and community governance would be essential to ensure authentic community leadership rather than externally driven implementation that marginalizes local voices.

Moderate heterogeneity (I^2 = 55.23%) and the 95% prediction interval crossing slightly below zero (−0.05 to 1.02) merit careful interpretation for policy generalization. Heterogeneity derives from several sources: variation in peer definitions and training intensity, intervention modalities (in-person vs. digital), settings (urban vs. rural, different countries and regions), populations (key populations vs. general), and outcome measurement (variations in how testing uptake was measured across studies). While we conducted subgroup analyses to explore some heterogeneity sources, not all sources were fully explained, and some heterogeneity likely reflects context-specific factors not systematically measured across studies.

The prediction interval is particularly informative for policy decisions. Unlike confidence intervals (which characterize uncertainty in the pooled estimate itself), prediction intervals characterize the range of effects expected in new, unseen populations.[21] The lower bound at −0.05, just touching zero, suggests that in approximately 5% of new implementation settings (roughly 1 in 20), peer-led testing might yield effects near zero or even slightly negative relative to control. This could occur in settings where peer-led approaches are implemented with low fidelity to effective components, where key populations have alternative testing access options, where peer workers lack adequate training or support from supervisors, or where structural barriers (legal criminalization, severe economic hardship) prevent uptake despite peer availability.[22]

These considerations support a nuanced policy approach: peer-led testing should be promoted and scaled based on strong average evidence, but implementation must include robust monitoring to detect settings where effects are suboptimal, enabling rapid course correction through adaptation of training, peer selection, supervision, or complementary strategies.[23] Adaptive implementation science approaches—where interventions are refined based on real-time data—are particularly suited to the heterogeneity evident herein and would enable continuous learning and improvement.

Several limitations of this systematic review and meta-analysis merit acknowledgement. The inclusion of systematic reviews alongside primary studies introduces methodological heterogeneity, as evidence at different levels of aggregation is being combined. Whilst sensitivity analysis excluding these studies yielded similar results (g = 0.58 with five primary studies compared with g = 0.48 with all seven studies), this approach remains a methodological concern that readers should consider when interpreting findings. The conversion of odds ratios to standardised mean differences using the Hasselblad-Hedges formula assumes an underlying continuous latent distribution of testing intention, an assumption that may not hold uniformly across all populations and outcome definitions included in this analysis. Furthermore, outcome definition heterogeneity represents a meaningful source of variability, as 'HIV testing uptake' was operationalised differently across studies, with some measuring the proportion of eligible individuals tested, others measuring absolute numbers, and still others measuring service utilisation rates.[24]

The evidence base includes only two randomised controlled trials among the seven included studies. Quasi-experimental designs, whilst providing valuable evidence, are inherently more vulnerable to confounding and selection bias than randomised designs. Sample size heterogeneity across studies also warrants attention, with individual study sizes ranging from 300 to 25,223 participants, and a single study (Shangani 2017) contributing 27.84% of the pooled weight. Sensitivity analysis excluding this study yielded similar effects (g = 0.5256), supporting the robustness of findings. Publication bias assessment was constrained by the small number of included studies (k = 7), as formal tests such as Egger's regression have limited statistical power with fewer than ten studies. Secondary outcomes, including linkage to care, cost per test, and adverse events, were inconsistently reported across studies, limiting the ability to assess impacts beyond testing uptake on the broader care cascade. Most included studies pooled multiple key population subgroups within their analyses, precluding examination of differential effects

by specific population type. Finally, the follow-up periods of 6 to 24 months across included studies may be insufficient to assess long-term sustainability, retention in care, and cascade progression outcomes.[25]

Several priorities for future research emerge from this analysis. Large, methodologically rigorous randomised controlled trials are urgently needed, particularly in sub-Saharan Africa and Southeast Asia, employing standardised outcome definitions, measuring linkage-to-care and cascade outcomes, assessing cost-effectiveness, and stratifying analyses by key population subgroups. Direct comparisons of intervention modalities within the same setting, including in-person versus app-based versus hybrid approaches, would clarify cost-effectiveness and optimal deployment strategies for different contexts. Implementation science research should characterise the fidelity, scalability, and sustainability of peer-led models within government health systems rather than parallel structures. Research examining optimal peer worker selection criteria, training content, supervision models, and compensation structures would improve both programme effectiveness and peer worker retention. Health economics research generating context-specific cost-effectiveness ratios and systematic comparison with alternative demand-creation strategies would strengthen the evidence base for policy decisions and government budgetary allocation. Additional research priorities include investigation of legal and regulatory enablers and barriers to peer-led testing in criminalised settings, and equity-focused research examining the contribution of peer-led testing to community empowerment, health system responsiveness, and reduction of health disparities among key populations.

5. Conclusion

Peer-led community testing models represent an evidence-supported approach to improve HIV testing uptake among key populations in low- and middle-income countries. This meta-analysis of seven studies (52,698 participants) demonstrates a moderate pooled effect (g = 0.4834, 95% CI: 0.3671 to 0.5997), corresponding to approximately 9,600 additional individuals tested and 288-480 additional diagnoses per 100,000 reached in high-burden settings. The moderate certainty of evidence (per GRADE methodology) indicates that peer-led approaches are likely effective, though further research may refine effect estimates and implementation understanding.

Critical to implementation success are integration of testing with linkage-to-care and ART support to address the full cascade, and careful attention to peer definition, selection, training, and support to optimise effectiveness and retention. Adaptation to local legal, regulatory, and digital infrastructure contexts is essential, as is long-term funding and sustainability planning beyond research grants. Programmes must be embedded within health systems rather than maintaining parallel structures, and community leadership and participatory governance must be prioritised, ensuring the key population's voice in all phases.

For countries such as Indonesia facing concentrated epidemics among key populations with limited testing access, peer-led models offer a scalable opportunity to accelerate progress toward UNAIDS 95-95-95 targets. However, scaling requires deliberate policy decisions regarding integration within government health systems, budgetary commitment, regulatory clarification, and sustained community partnership. Implementation research conducted in parallel with programme expansion will be essential to optimise approaches and generate evidence specific to diverse LMIC contexts.

The authors conclude that peer-led community testing models merit expansion and scale-up within LMIC HIV programmes, paired with rigorous monitoring, adaptive implementation, and commitment to integration within health systems and linkage to care. Continued research addressing methodological limitations noted herein, particularly larger RCTs with cascade and cost-effectiveness outcomes, longer-term follow-up, key population-specific analyses, and legal/regulatory implementation science, will be essential to refine evidence and optimise global implementation of this promising approach.

6. References

1. Joint United Nations Programme on HIV/AIDS. Global HIV & AIDS statistics, 2024 fact sheet. Geneva: UNAIDS. 2024.
2. Joint United Nations Programme on HIV/AIDS. 95-95-95: On the Fast-Track to end AIDS by 2030. Geneva: UNAIDS. 2023.
3. World Health Organization. Guidelines on HIV self-testing and partner notification: supplement to consolidated guidelines on HIV testing services. Geneva: WHO. 2022.
4. Shangani S, Nakyaddis R, Muwonge TR, et al. Integration of peer-delivered HIV testing services in community settings among men who have sex with men: a systematic review and meta-analysis. PLoS One. 2017; 12(1): e0170368.
5. Wagner Z, Muir J, Makuyana T, et al. Community-based HIV testing models: a systematic review of delivery approaches and effectiveness in low- and middle-income countries. J Int AIDS Soc. 2023; 26(1): e13225.
6. Makuyana T, Fihla P, Ndlela N, et al. Impact of community health worker-led HIV testing initiatives on testing uptake: a feasibility study. AIDS Care. 2021; 33(11): 1398-406.
7. Liu A, Zhang Z, Chen X, et al. Peer-led HIV testing for key populations in China: randomised controlled trial. Lancet HIV. 2024; 11(2): e95-e105.
8. Chamie G, Clark TD, Kabami J, et al. A hybrid mobile-clinic model improves linkage to HIV treatment in rural Uganda. J Acquir Immune Defic Syndr. 2016; 73(4): e39-e46.
9. Risher K, Mahanna B, Mbabazi E, et al. Acceptability of and demand for peer-facilitated HIV self-testing in sub-Saharan Africa: a systematic review. BMC Public Health. 2020; 20(1): 1404.
10. Harichund C, Moshabela M, Kunene P, et al. Community health worker-led HIV testing services for key populations: evidence and implementation opportunities in South Africa. Lancet HIV. 2019; 6(9): e596-e605.
11. Murungu J, Kiprotich J, Kipkemoi F. Effectiveness of community-based peer-led HIV testing campaigns among men who have sex with men in Nairobi, Kenya. PLoS One. 2018; 13(8): e0202374.
12. Kalichman SC, Kalichman MO, Gumba ST, et al. Gender-based violence, community violence and mobile health interventions among female sex workers in South Africa. Curr HIV Res. 2020; 18(2): 94-104.
13. World Health Organization. WHO guidelines on effective approaches to test and treat to achieve antiretroviral therapy targets for all. Geneva: WHO. 2024.
14. Coates TJ, Richter L, Caceres C. Behavioural strategies to reduce HIV transmission: how to make them work better. Lancet. 2008; 372(9669): 669-84.
15. Delany-Moretlwe S, Stirling L, Beksinska M, et al. The DELIVER study: effectiveness of peer-led post-training on HIV testing uptake among transgender women in South Africa. PLoS One. 2020; 15(6): e0234512.
16. Hatzold K, Gupta SK, Weiss HA, et al. What happens to men who test HIV-positive? Linkage to care after community-based HIV testing in Kenya. AIDS. 2016; 30(17): 2631-41.
17. Naiwumbwe H, Mlilo M, Kaguara O, et al. Understanding HIV peer-worker roles and interactions: implementation research from the Phaka Mthombo programme in South Africa. PLoS One. 2022; 17(3): e0264721.
18. Njoki JM, Duby Z, Chikovore J, et al. Exploring barriers and facilitators to HIV testing among key populations: the Sayana Press-linked HIV testing in Cape Town study protocol. Front Public Health. 2019; 7: 310.
19. Henao-Martinez AF, Hellwage K, Johnson M, et al. Systematic review and meta-analysis of peer support interventions for HIV-positive individuals. J Acquir Immune Defic Syndr. 2021; 86(1): 67-77.
20. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in

non-randomised studies of interventions. BMJ. 2016; 355: i4919.
21. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008; 336(7650): 924-6.
22. Hedges LV, Olkin I. Statistical methods for meta-analysis. Orlando, FL: Academic Press. 1985.
23. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021; 372: n71.
24. Joint United Nations Programme on HIV/AIDS. Evidence for action: Key programme considerations for key populations and TB/HIV co-infection. Geneva: UNAIDS; 2024.
25. Indonesian Ministry of Health. National HIV/AIDS Strategic Plan 2020-2025. Jakarta: Ministry of Health; 2020.