Evaluating HIV prevention effectiveness: the perfect as the enemy of the good

Marie Laga[a], Deborah Rugg[b], Greet Peersman[c] and Martha Ainsworth[d]

There is a need to better understand the effectiveness of HIV-prevention programs. Cluster randomized designs have major limitations to evaluate such complex large-scale combination programs. To close the prevention evaluation gap, alternative evaluation designs are needed, but also better articulation of the program impact pathways and proper documentation of program implementation. Building a plausible case using mixed methods and modeling can provide a valid alternative to probability evidence. HIV prevention policies should not be limited to evidences from randomized designs only.

AIDS 2012, **26**:779–783

Keywords: HIV prevention, impact evaluation, plausibility, randomized controlled trial

The HIV prevention evaluation gap

Intensifying HIV prevention is the only way to defeat the epidemic ultimately, and essential to respond to the growing unmet treatment needs [1–3].

With no vaccine to block transmission, HIV prevention focuses on reducing transmissibility and risk [4]. Prevention programs involve multiple approaches aiming at reducing risky behavior, promoting uptake of and adherence to essential prevention tools such as condoms, clean needles or male circumcision, as well as empowering communities and creating an enabling environment [4,5]. This mix of biomedical, behavioural and structural interventions now referred to as 'Combination Prevention' offers the best promise of success [6]. Evaluating the impact of those combination programs on lowering HIV incidence at the population level remains challenging.

Evidence that HIV prevention can work has been accumulating from country experiences such as Thailand [7] and evaluations of well defined, mostly biomedical, components of prevention programs [8–11]. The most recent breakthrough showed that early antiretroviral therapy (ART) could reduce transmission by 97% in discordant couples [12]. The potential impact of 'early treatment as prevention' at the population level in different settings is still unknown, and behavioral and community approaches will be needed to implement it effectively. So far, data on the impact of combination prevention programs are scarce, and cluster randomized controlled trials (c-RCT) assessing impact of preventive interventions on HIV incidence have not demonstrated an effect [13–19]. As a result, we are left with a gap in understanding 'what works' in HIV prevention, attributed to a lack of evaluation culture or poor prevention science [8–11]. We argue that the methodological

[a]Department of Public Health, Institute of Tropical Medicine, Antwerp, Belgium, [b]Inspection and Evaluation Division, UN, OIOS, New York, [c]Payson Center for International Development, Tulane University, New Orleans, and [d]Independent Evaluation Group, World Bank, Washington, District of Columbia, USA.

Correspondence to Marie Laga, Department of Public Health, Institute of Tropical Medicine, Nationalestraat 155, 2000 Antwerp, Belgium.

Tel: +32 3 24763 16; fax: +32 3 24765 32; e-mail: mlaga@itg.be
Received: 13 October 2011; revised: 5 January 2012; accepted: 27 January 2012.

DOI:10.1097/QAD.0b013e328351e7fb

challenges in measuring the effectiveness of combination HIV prevention programs should not be underestimated, and advocate for realism and pragmatism when it comes to generating more convincing evidence to guide prevention programming.

This article is based in part on discussions held at a Joint United Nations Programme on HIV/AIDS (UNAIDS) Think Tank meeting on HIV prevention evaluation methodologies in Sussex, England in 2009 [20].

Shortcomings of the 'gold standard' in evaluating HIV prevention programs

The RCT is the gold standard for the evaluation of drugs or biomedical prevention tools. Randomized designs with the community as unit of intervention (c-RCT) have been increasingly used for impact evaluations of public health and development programs with mixed success generating controversy [21–24]. Behavioural scientists questioned the 'appropriateness' of experimental designs to evaluate effectiveness of HIV health promotion since the 1990s [25,26]. Whether the c-RCT with HIV incidence as the impact indicator should become the gold standard to determine what works in HIV prevention programming is currently a matter of fierce debate [27,28]. A recent call for funding of combination prevention evaluation restricted funding proposals to 'randomized designs only' [29]. We argue that it is a missed opportunity to restrict the scarce program evaluation resources to highly expensive c-RCT, which may not bring valid answers. For a number of reasons, those designs will likely never fill the evidence gap.

Challenges in measuring change in HIV incidence

The first obstacle is the lack of reliable, easy to use tools to measure HIV incidence at a population level. Direct estimation through follow-up of a cohort is complex, costly, and unsustainable outside of research settings. Laboratory assays have been developed to estimate incidence but still face problems with validity [30]. Work is ongoing to improve and develop new assays [31,32]. Incidence estimated indirectly by relying on proxy methods [33] or mathematical models [34,35] can contribute to build a plausible case of prevention success, if triangulated with other program data.

A second challenge is that HIV incidence is low in most populations, requiring unrealistically large sample sizes [2]. If direct HIV incidence measurement is held up as an absolute requirement, our ability to assess prevention interventions will continue to be sharply curtailed.

Intermediate indicators, such as reported behavior change or sexually transmitted infection (STI) rates, have been used as proxies for impact, because they are easier to measure, and more prevalent. But there are issues of reliability with reported sexual behavior and variable links between STI and HIV [36,37]. Trends in STI incidence or reported behavior change may not provide definite proof of impact, but are important data to triangulate with other evidence to build a plausible case of prevention success.

The limitations of randomized designs

The theoretical advantages of c-RCT are clear: high level of rigor, high internal validity, and quantifiable level of evidence that allows for estimating cost-effectiveness [20]. The main shortcomings are high complexity with large sample size and cost, difficulty to find 'naive' control communities or ethical concerns about withholding interventions to the control group [21,38]. And because success of HIV prevention programs depend heavily on the context and the epidemic stage, external validity is likely to be low [39].

Intriguingly, so far none of the seven published c-RCT evaluating 'combination' prevention programs has found an impact on HIV incidence [13–19]. Different explanations for the flat results have been suggested, varying from the interventions were truly not effective [28] to all possible factors related to the design of those c-RCT [13–19]. An additional issue is that potentially important program components of combination prevention may never be evaluated because they are not amenable to experimental designs. As the authors of the community RCT of multicomponent adolescent sexual health program Mema Kwa Vijana [14] stated 'Interventions were deliberately constrained to be affordable and replicable on a large scale. The trials design also meant that mass media approaches and region-wide approaches could not be included.' This 'fitting the intervention to the trial' driven by need for rigor is particularly worrisome for HIV prevention evaluation, as it restricts evaluations to well defined, mostly biomedical interventions. The behavioral, social and contextual components, as essential elements or strong enablers, are left out or kept to a minimum. As a result, the evidence base for potentially stronger multicomponent programs remain poor, not because they do not work but because they are not suited to be evaluated in a c-RCT design [40].

In the end, the flat results of those well conducted trials are difficult to interpret, and created confusion for program planners and policy makers. Absence of evidence has been confused with absence of effectiveness.

Closing the prevention evaluation gap

To close the prevention evaluation gap, alternative evaluation designs are needed, but also better articulation

of the program impact pathways (PIPs) and proper documentation of program implementation.

Alternative evaluation designs

In other fields of public health, it has been recognized that the probability design, such as c-RCT, is often inappropriate for impact evaluation of large scale programs, and plausibility designs have been proposed as a valid alternative [38]. Plausibility designs provide a lower level of certainty in linking the program to any observed changes than probability designs, as the control groups are not randomly selected. The aim is to build a plausible case of program impact, which is shown by convergence of evidence using triangulation of different data sources. Avahan, a large-scale sex worker program in India, built a plausible case of the program's impact on HIV incidence by combining data from program monitoring, process evaluation, consecutive population-based surveys, qualitative methods and modeling [41,42]. Counterfactuals for program effects were nonrandom and included before and after analysis, comparisons with nonintervention areas and modeled 'control areas'. Quasiexperimental designs using 'statistical' control groups instead of random assignment have been increasingly across multiple fields of programme evaluation [43].

Building a plausibility case of evidence from mixed methods can give more meaningful and relevant insights for program managers and policy makers because it addresses not only the question 'whether the program had impact' but also provides information about 'how the effect was obtained or why the program was effective'.

Interpreting national HIV trends can also provide insights into prevention effectiveness. In the absence of a 'control' group to compare results with, it is challenging to interpret those trends. Assessing whether the expected changes occurred is referred to as 'Adequacy' design by Habicht *et al*. [21]. The effect of national prevention efforts on HIV trends has been assessed retrospectively by linking them credibly to the prevention activities at the time and subsequent behavior change. In Zimbabwe, a 50% decline in HIV prevalence was observed between 1997 and 2007, and through triangulation of different data sources, partner reduction emerged as an important factor that contributed to the prevention success [44].

Articulating how the program will reduce HIV transmission

Prevention programs typically have a long causal pathway and are inherently complex because they involve multiple groups and approaches that directly or indirectly impact on HIV transmission at the population level.

It is important to clearly describe the program components and how they are intended to make a difference on HIV transmission, laying out the nature of the causal pathway and making intermediate outputs and outcomes explicit. Each step of a PIP should as much as possible be based on known theories and mechanisms of change, supported by available evidence. Program components all need to be part of one of the direct or indirect links leading to HIV incidence reduction. Constructing a clear PIP is not just good practice in program design but provides the basis for strong evaluation [35]. It sets the stage for asking not only whether a preventive intervention or program worked in lowering HIV transmission, but understanding how and why it worked.

Unfortunately, good examples of PIPs of HIV prevention programs are rare, or at least unavailable in the accessible literature. Learning from other similar fields is key here. 'Intervention mapping' has been a useful program planning tool developed by health promotion specialists allowing to take into account theory, evidence and context [45].

'Complexity' has been an important challenge in HIV prevention programming and evaluation. It has led to 'magic bullet thinking', reflected in a tendency to prioritize the well defined biomedical interventions leaving out the less definable social and contextual approaches. The goal of a better articulation of the program and it's impact pathways is to assist planners and evaluators in simplifying the complex reality without becoming simplistic.

Monitoring of implementation and uptake of prevention programs

In order for a program to make a difference in the desired outcome, it needs to be appropriate to the specific context and implemented at sufficient scale, coverage and quality. Too often, evaluators try to measure impact without documenting first the program coverage, uptake and intensity, which are key determinants of the program's success. The weak program monitoring in HIV prevention to date is evident [2,12]. Most national response analyses are limited to a list of poorly defined prevention interventions with little or no indication about scale, reach and coverage of the group targeted or of the quality of implementation. Either the data are not collected, the data are not made available to the national AIDS coordination body or the data have not been analyzed or used. This is in stark contrast with ART program monitoring. Proof that it can be done is AVAHAN, which has been exemplary in program monitoring at all levels [46].

There are still challenges involved in measuring coverage, estimating the size of hard-to-reach populations, defining and monitoring minimum quality standards in HIV prevention, as well as in monitoring changes in social and structural determinants [47]. Progress has been made recently with regard to specific methodologies to quantify

hard to reach, hidden or highly stigmatized populations [48,49]. However, with the already available tools and methods, countries could do much better in documenting implementation of their prevention programs.

Conclusion

The need to better document the effectiveness of HIV prevention programs and to do this with the most robust methods possible is widely recognized. We argue that by limiting prevention program evaluation to experimental methods and HIV incidence as outcome, the perfect becomes the enemy of the good. The evidence base of 'what works in prevention, where and for whom?' will remain incomplete, sustaining confusion for program planners and contributing to the crisis of confidence in combination prevention, and subsequent inaction.

We have made concrete suggestions on how to move forward in terms of improving prevention evaluation.

First, we need to be more flexible and adaptive in choosing methods to evaluate prevention effectiveness. Building a plausible case using mixed methods, convergence of data sources, and modeling gets us a long way in the evaluation of combination prevention programs. They can provide a valid alternative to probability evidence and may be more persuasive in terms of why and how interventions work in different contexts.

Second, an explicit PIP should become standard practice in prevention programming, not only to improve planning but also as a useful framework for monitoring and evaluation activities.

Third, program managers need to integrate monitoring and evaluation strategies into their programs from the start, and collect relevant information if they want to learn while doing. This requires a closer collaboration between implementers, evaluators and policy makers, and a willingness of donors to fund those activities.

Fourth, there is a clear need to develop incidence assays that give reliable HIV incidence measures from cross-sectional surveys. In the meantime, modeling can produce helpful proxy incidence estimates for impact evaluation.

Experimental designs will continue to have a place in the evaluation of specific well defined components of prevention programs. But with two million new HIV infections a year, we cannot afford to dismiss potentially effective prevention programs simply because they can't easily be randomized or because they are 'too complex' to evaluate. Evidence-based HIV prevention is possible, but it must go beyond RCTs.

Acknowledgements

All four authors played an active role in the writing of the different versions of the article. M.L. is the corresponding author.

The findings, interpretations, and conclusions expressed in this article are entirely those of the authors and should not be attributed in any manner to the World Bank, its Board of Executive Directors, or the governments they represent.

Conflicts of interest

There are no conflicts of interest.

References

1. Sidibe M, Buse K. **Fomenting a prevention revolution for HIV.** *Lancet* 2010; **375**:533–535.
2. UNAIDS. UNAIDS Global Report, (2010). http://www.unaids.org. [Accessed 6 February 2012].
3. Van Damme W, Kober K, Laga M. **The real challenges for scaling up ART in sub-Saharan Africa.** *AIDS* 2006; **20**:653–656.
4. UNAIDS. Practical guidelines for intensifying HIV prevention. Geneva, Switzerland: Joint United Nations Programme on HIV/AIDS, (2007). http://www.unaids.org. [Accessed 6 February 2012].
5. Schwartländer B, Stover J, Hallett T, Atun R, Avila C, Gouws E, *et al*. **Toward an improved investment approach for an effective response to HIV/AIDS.** *Lancet* 2011; **377**:2031–2041.
6. Hankins CA, De Zalduondo BO. **Combination prevention: a deeper understanding of effective HIV prevention.** *AIDS* 2010; **24 (Suppl4)**:S70–S80.
7. Celentano DD, Nelson KE, Lyles CM, Beyrer C, Eiumtrakul S, Go VF, *et al*. **Decreasing incidence of HIV and sexually transmitted diseases in young Thai men: evidence for success of the HIV/AIDS control and prevention program.** *AIDS* 1998; **12**:F29–F36.
8. Padian NS, Buvé A, Balkus J, Serwadda D, Cates W Jr. **Biomedical interventions to prevent HIV infection: evidence, challenges, and way forward.** *Lancet* 2008; **372**:585–599.
9. Coates TJ, Richter L, Caceres C. **Behavioural strategies to reduce HIV transmission: how to make them work better.** *Lancet* 2008; **372**:669–684.
10. Gupta GR, Parkhurst JO, Ogden JA, Aggleton P, Mahal A. **Structural approaches to HIV prevention.** *Lancet* 2008; **372**:764–775.
11. Bertozzi SM, Laga M, Bautista-Arredondo S, Coutinho A. **Making HIV prevention programmes work.** *Lancet* 2008; **372**:831–844.
12. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, *et al*. **Prevention of HIV-1 infection with early antiretroviral therapy.** *N Engl J Med* 2011; **365**:493–505.
13. Kamali A, Quigley M, Nakiyingi J, Kinsman J, Kengeya-Kayondo J, Gopal R, *et al*. **Syndromic management of sexually-transmitted infections and behaviour change interventions on transmission of HIV-1 in rural Uganda: a community randomised trial.** *Lancet* 2003; **361**:645–652.
14. Ross DA, Changalucha J, Obasi AI, Todd J, Plummer ML, Cleophas-Mazige B, *et al*. **Biological and behavioural impact of an adolescent sexual health intervention in Tanzania: a community-randomized trial.** *AIDS* 2007; **21**:1943–1955.
15. Corbett EL, Makamure B, Cheung YB, Dauya E, Matambo R, Bandason T, *et al*. **HIV incidence during a cluster-randomized trial of two strategies providing voluntary counselling and testing at the workplace, Zimbabwe.** *AIDS* 2007; **21**:483–489.
16. Jewkes R, Nduna M, Levin J, Jama N, Dunkle K, Puren A, Duvvury N. **Impact of stepping stones on incidence of HIV and HSV-2 and sexual behaviour in rural South Africa: cluster randomised controlled trial.** *BMJ* 2008; **337**:a506.

17. Cowan FM, Pascoe SJS, Langhaug LF, Mavhu W, Chidiya S, Jaffar S, *et al*. **The Regai Dzive Shiri Project: results of a randomized trial of an HIV prevention intervention for youth.** *AIDS* 2010; **24**:2541–2552.
18. Pronyk PM, Hargreaves JR, Kim JC, Morison LA, Phetla G, Watts C, *et al*. **Effect of a structural intervention for the prevention of intimate-partner violence and HIV in rural South Africa: a cluster randomised trial.** *Lancet* 2006; **368**:1973–1983.
19. Gregson S, Adamson S, Papaya S, Mundondo J, Nyamukapa CA, Mason PR, *et al*. **Impact and process evaluation of integrated community and clinic-based HIV-1 control: a cluster-randomised trial in eastern Zimbabwe.** *PLoS Med* 2007; **4**:e102.
20. UNAIDS. Strategic guidance for evaluating HIV prevention programmes. Geneva, Switzerland: Joint United Nations Programme on HIV/AIDS, (2010). http://www.unaids.org. [Accessed 6 February 2012].
21. Habicht JP, Victora CG, Vaughan JP. **Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact.** *Int J Epidemiol* 1999; **28**:10–18.
22. Ravallion M. Should the randomistas rule? Economists' voice. (2009). Berkeley Electronic Press. http://www.bepress.com/ev. [Accessed 6 February 2012].
23. Ravallion M. Evaluating anti-poverty programs. In: Schultz TP, Strauss J, editors. *Handbook of development economics*. Amsterdam: North Holland; 2007.
24. Nutbeam D. **Evaluating health promotion, progress, problems and solutions.** *Health Promot Int* 1998; **13**:27–43.
25. Kippax S, Van de Ven P. **An epidemic of orthodoxy? Design and methodology in the evaluation of the effectiveness of HIV health promotion.** *Crit Public Health* 1998; **8**:371–386.
26. Van de Ven P, Aggleton P. **What constitutes evidence in HIV/aids education?** *Health Educ Res* 1999; **14**:461–471.
27. Padian NS, Holmes CB, McCoy SI, Lyerla R, Bouey PD, Goosby EP. **Implementation Science for the US President's Emergency Plan for AIDS Relief (PEPFAR).** *J Acquir Immune Defic Syndr* 2011; **56**:199–203.
28. Padian NS, McCoy SI, Balkus JE, Wasserheit JN. **Weighing the gold in the gold standard: challenges in HIV prevention research.** *AIDS* 2010; **24**:621–635.
29. 'Impact Evaluation of Combination HIV Prevention Interventions in PEPFAR Countries', CDC funding opportunity, RFA-GH-11-006. http://www.cdc.gov/od/pgo/funding/grants . [Accessed 22 December 2011].
30. Parekh BS, Kennedy MS, Dobbs T, Pau CP, Byers R, Green T, *et al*. **Quantitative detection of increasing HIV type 1 antibodies after seroconversion: a simple assay for detecting recent HIV infection and estimating incidence.** *AIDS Res Hum Retroviruses* 2002; **18**:295–307.
31. Hallett TB, Ghys P, Bärnighausen T, Yan P, Garnett GP. **Errors in 'BED'-Derived Estimates of HIV Incidence will vary by place, time and age.** *Plos ONE* 2009; **4**:e5720; doi:10.1371/journal.pone.0005720.
32. Incidence Assay Critical Path Working Group. **More and better information to tackle HIV epidemics: towards improved HIV incidence assays.** *Plos Medicine* 2011; **8**:e1001045.
33. Ghys PD, Kufa E, George MV. **Measuring trends in prevalence and incidence of HIV infection in countries with generalised epidemics.** *Sex Transm Infect* 2006; **82 (Suppl1)**:i52–i56; doi:10.1136/sti.2005.016428.
34. Ghys PD, Brown T, Grassly NC, Garnett G, Stanecki KA, Stover J, Walker N. **The UNAIDS Estimation and Projection Package: a software package to estimate and project national HIV epidemics.** *Sex Transm Inf* 2004; **80 (Suppl 1)**:i5–i9.
35. Hallett TB, Zaba B, Todd J, Lopman B, Mwita W, Biraro S, *et al*. **Estimating incidence from prevalence in generalised epidemics.** *Plos Med* 2008; **5**:e80.
36. Peterman TA, Lin LS, Newman DR, Kamb ML, Bolan G, Zenilman J, *et al*. **Does measured behavior reflect STD risk? An analysis of data from a randomized controlled behavioral intervention study. Project RESPECT Study Group.** *Sex Transm Dis* 2000; **8**:446–451.
37. Aral SO, Peterman TA. **A stratified approach to untangling the behavioral/biomedical outcomes conundrum.** *Sex Transm Dis* 2002; **9**:530–532.
38. Victora CG, Habicht JP, Bryce J. **Evidence-based public health: moving beyond randomized trials.** *Am J Public Health* 2004; **3**:400–40540.
39. Victora CG, Schellenberg JA, Huicho L, Amaral J, El Arifeen S, Pariyo G, *et al*. **Context matters: interpreting impact findings in child survival evaluations.** *Health Policy Plan* 2005; **20 (Suppl 1)**:i18–i31.
40. Kemm J. **The limitations of 'evidence-based' public health.** *J Eval Clin Pract* 2006; **12**:319–324.
41. Chandrasekaran P, Dallabetta G, Loo V, Mills S, Saidel T, Adhikary R, *et al*. **Evaluation Design For Large-Scale Hiv Prevention Programmes: the case of AVAHAN, the India AIDS initiative.** *AIDS* 2008; **22**:S1–S15.
42. Boily MC, Pickles M, Vickerman P, Buzdugan R, Isac S, Deering KN, *et al*. **Using mathematical modelling to investigate the plausibility of attributing observed antenatal clinic declines to a female sex worker intervention in Karnataka state, India.** *AIDS* 2008; **22 (Suppl 5)**:S149–S164.
43. Shadish WR, Cook TD, Campbell DT. *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton-Mifflin, 2002.
44. Halperin DT, Mugurungi O, Hallett TB, Muchini B, Campbell B, Magure T, et al. **A surprising prevention success: why did the HIV epidemic decline in Zimbabwe?** *Plos Med* 2011; **8**:e1000414.
45. Bartholomew LK, Parcel GS, Kok G, Gottlieb NH, Fernández ME. *Planning health promotion programs: an intervention mapping approach*, 3rd ed. San Francisco: Jossey Bass; 2011.
46. Verma R, Shekhar A, Khobragade S, Adhikary R, George B, Ramesh BM, *et al*. **Scale-up and coverage of AVAHAN: a large-scale HIV-prevention programme among female sex workers and men who have sex with men in four Indian states.** *Sex Transm Infect* 2010; **86**(Suppl 1):i76–i82.
47. Auerbach JD, Parkhurst JO, Caceres CF, Keller KE. **Addressing social drivers of HIV/AIDS: some conceptual, methodological and evidentiary considerations.** *AIDS* 2031; Working paper no. 24.
48. UNAIDS/WHO. Guidelines on estimating the size of populations most at risk to HIV. UNAIDS/WHO Working group on global HIV/AIDS and STI Surveillance. http://www.who.int. [Accessed 6 February 2012].
49. Vuylsteke B, Vandenhoudt H, Langat L, Semde G, Menten J, Odongo F, *et al*. **Capture-recapture for estimating the size of the female sex worker population in three cities in Côte d'Ivoire and in Kisumu, western Kenya.** *Trop Med Int Health* 2010; **15**:1537–1543.