Appropriate evaluation of HIV prevention interventions: from experiment to full-scale implementation

Timothy B Hallett, Peter J White, Geoff P Garnett

Background: Preventing HIV infection is still an essential goal in tackling the HIV/AIDS pandemic. Remarkably little is known about how best to reduce HIV incidence because most trials focus on the reduction of risk behaviours and assume an effect on HIV incidence.

Objective: To discuss the evidence for the effectiveness of HIV prevention strategies, exploring the different types of evidence available: individual and community randomised controlled trials, and observational studies.

Results: Although providing a gold standard for evidence, trials have been limited in their scope and are difficult to interpret and generalise. There have been examples of national level successes in preventing HIV which have been detected in surveillance data and understood through behavioural and modelling studies. These have the advantage of being to scale and indicating effectiveness rather than efficacy.

Conclusions: Although randomised trials are important because of their scientific rigor, it is also important that evidence from observational epidemiology is not overlooked. Only if good quality, consistent data are available can the history of the HIV epidemic be appropriately analysed.

Substantial resources have been committed globally to the treatment and prevention of HIV and AIDS and advances are being made in providing antiretroviral treatments in resource-poor settings. Such advances will be unsustainable without successes in HIV prevention. However, remarkably little is known about the effectiveness of preventive interventions in reducing the incidence of HIV itself. In part, the “emergency” nature of the HIV epidemic has required implementation of interventions to run ahead of evidence of effectiveness. In planning a response to the global pandemic assumptions have been made about what works, using limited evidence from theory, the measurement of processes or intermediate intervention outcomes, such as behaviour change.

Here we discuss how interventions have been evaluated and interpreted and the epidemiological and methodological issues that have been learned. National surveillance, with behavioural data collection as an adjunct, can provide the most powerful and relevant test of HIV interventions.

INDIVIDUAL RANDOMISED CONTROLLED TRIALS

The randomised controlled trial (RCT) provides the “gold standard” for testing the efficacy of medical interventions, providing a clear experimental design to demonstrate the presence or absence of an effect. This study design works well when there is a straightforward replicable intervention linked to the disease outcome.

The majority of RCTs in research on the prevention of HIV use outcome measures of risk behaviour rather than HIV incidence itself. This is appropriate for interventions aimed at HIV-positive individuals, which show decreased unprotected sex and sexually transmitted infections (STIs), and when the STIs are themselves the outcome of interest. It is less clear how useful intermediate outcomes are in other situations exploring the risks of HIV either in the individual or in the community. For example, a recent measure of an HIV intervention in a community in Uganda found that condom use increased and numbers of sex partners decreased.

Many HIV intervention trials have now been performed and are summarised in systematic reviews, but the focus has been on how to change risk behaviour, with a concentration in the USA and on men who have sex with men, and less emphasis on populations in developing countries or on HIV incidence as an outcome.

There are many reasons why interventions to prevent the spread of HIV are not always readily assessed in individual RCTs.

- A key problem is the low incidence of HIV in many industrialised countries where intervention trials are well resourced. The adoption of intermediate outcomes such as increased condom use, reduced numbers of sexual partners or reduced incidence of other STIs is common. Data on self-reported behaviour change may be unreliable. Even if the changes are genuine it is not clear that such changes will actually alter HIV incidence depending on how they do on the epidemiological context. Changes in risk behaviour could be overwhelmed by high levels of exposure or could differentially influence lower risk contacts. Even where HIV incidence is sufficiently high to measure without prohibitively large sample sizes, the complexity and cost of measuring incident HIV infection has led to the use of proxy outcomes.

- Interventions aimed at HIV prevention are often complex, requiring skilled staff to carry out counselling and education to motivate and facilitate the adoption of safer sexual behaviours. Difficulties in ensuring that interventions are appropriately implemented create uncertainty about negative trial results and concerns about replicating the intervention outside the context of the trial.

- Other HIV interventions are not amenable to individual randomisation if the focus is the group, community or population—for example, the use of mass media advertising or educational drama can only be delivered to groups.

Conceptually, the focus on individuals may not be appropriate because individual behaviour is greatly affected by social and cultural norms, making interventions targeted only at selected individuals largely ineffective. This is particularly the case with sexual behaviour, which must

Abbreviations: CRCT, community RCT; RCT, randomised controlled trial; STI, sexually transmitted infection
study biases caused by poor follow-up rates, migration and extended follow-up period thus required increases the risks of population (Hallett one group can take time to percolate to the rest of the impact on incident HIV. Furthermore, the effect of changes in randomised. Because health promotion builds over time as aimed at preventing HIV. The complexity of HIV interventions unfortunately, the first three criteria are often absent in CRCTs (COMMUNITY RCTS)

The failure of individual-level trials to determine the community-level impact of interventions has led to a number of community RCTs (CRCTs) of HIV and STI interventions. In these, the intervention is targeted at communities which are randomised to treatment or control arms. Randomisation is used to distribute all the differences between the communities (known and unknown) evenly to both arms of the trial. The appropriate circumstances for a CRCT have been listed by Susser as requiring:

1. A well-defined, narrow hypothesis, so that the key to the intervention success or failure can be identified;
2. A measurable intervention, to be able to assess its implementation;
3. Adequate statistical power; and
4. Well-defined, measurable outcomes.

Unfortunately, the first three criteria are often absent in CRCTs aimed at preventing HIV. The complexity of HIV interventions and their delivery can be a problem for CRCTs just as it can for individual RCTs, but CRCTs also suffer from their scale, which presents major challenges of logistics and resources, often leading to compromises in the numbers of communities randomised. Because health promotion builds over time as knowledge increases and cultural norms and practices change, it is possible that it will take some time for the interventions to impact on incident HIV. Furthermore, the effect of changes in one group can take time to percolate to the rest of the population (Hallett et al, submitted for publication). The extended follow-up period thus required increases the risks of study biases caused by poor follow-up rates, migration and changes in the background epidemiological context, such as the influence of other STIs appears to be greatest in the early stages of an HIV epidemic when infection is concentrated in high-risk parts of the population and when the curable bacterial STIs are more important than infections such as genital herpes, which cannot be cured.

Interpreting positive CRCT results

A positive result of a CRCT conclusively demonstrates that the intervention has worked in a particular setting. However, questions remain:

- What are the necessary components of the intervention?
- What quality of intervention is required to maintain effectiveness in other settings?
- How well can the intervention be scaled up to a wider population or to a national level?

Essentially there is a problem translating the efficacy observed in trial conditions to effectiveness on a larger scale because interventions are often implemented less well when scaled-up (ie they can be less intensive, less carefully monitored and carried out in unselected populations with lower motivation). The costs of the intervention will probably be less outside the trial setting but at the same time this could compromise the effectiveness. In addition to problems of scale there is the question of whether the intervention will work in other epidemiological, social and cultural contexts. For example, the syndromic management of STIs significantly reduced HIV incidence in the Mwanza trial but this result has not been reproduced in the other trials of STI control targeting HIV incidence. The influence of other STIs appears to be greatest in the early stages of an HIV epidemic when infection is concentrated in high-risk parts of the population and when the curable bacterial STIs are more important than infections such as genital herpes, which cannot be cured.

Interpreting negative CRCT results

Negative CRCT results are even more difficult to interpret. They tell us that the intervention, as implemented, did not produce significant results, but there are a number of potential explanations. The study may have been inadequately powered because sample-size calculations require knowledge of the variation of incidence between the populations studied, and accurate forecasts of future patterns of incidence without interventions (Hallett et al, submitted for publication). Because of the dynamic nature of epidemics such future patterns may not be consistent with past incidence rates, and variation in recorded prevalence at baseline may be a misleading proxy of variation in incidence. Implementation of the intervention may not have been adequate; however, because effectiveness is more important than efficacy, and as within intervention trials the interventions are likely to be more assiduously implemented than when scaled-up, this may seem a minor problem. Nonetheless, the complex nature of many interventions means that a failure of implementation rather than efficacy can be claimed and failing interventions be continued. More reasonably, the generalisability of the negative result to other epidemiological contexts can be debated. For individual RCTs, repetition of the trial in the same and in other contexts can eliminate doubts through comparison and meta-analyses of multiple trials. Unfortunately, these are not easily achieved with long and costly CRCTs.
OBSERVATIONAL STUDIES

The urgency required in responding to the spread of a lethal infectious disease has led to many interventions being implemented before the evidence from trials has become available. In some cases where large-scale interventions have been implemented, associated observational epidemiological and behavioural studies have identified changes in HIV prevalence that can only be attributed to changes in HIV incidence and have been linked to changes in specific risk behaviours.7-29 Such observational studies have the advantage of assessing interventions at the regional or national scale. This immediately avoids many problems of trials, as the interventions are already to scale and measure effectiveness. However, in the hierarchy of evidence, such observational epidemiology is considered to be weak. Biases such as changes in the representativeness of antenatal clinics included in sentinel surveillance, changes in fertility due to HIV affecting data from antenatal clinics, trends unrelated to the interventions, and the natural dynamics of the incidence of STI and HIV can lead to spurious measures of success or failure. Nonetheless, if such biases can be predicted and measured their effects can be excluded and powerful arguments of what works can be based on observational epidemiological data.

In the late 1980s and early 1990s, in response to the high prevalence of HIV, the government of Uganda implemented an HIV prevention strategy acknowledging and promoting the discussion of HIV risks and advocating a reduction in sexual risk behaviour.30 The prevalence of infection was seen to decline in young women attending antenatal clinics. Cross-sectional population-based survey data showed decreases in reported numbers of sexual partners.25-32 Such declines, particularly in those age groups too young to have had HIV-infected individuals removed by death or subfertility, could only be explained if HIV incidence had declined and provided a much-needed national level success story. Subsequent debate has continued over the cause of the success, with the neologism ABC—abstinence, be faithful, use condoms—being introduced, and competing claims being made over the relative importance of condom use and abstinence.32-33 A similar national level change in patterns of prevalence and incidence was achieved in Thailand. Here, studies that found sex work was the major cause of this behaviour change occurs in year 1990, 1995 or 2000 (arrows). The analyses of HIV trends in Zimbabwe and Kenya were restricted to those antenatal clinics sampled continuously over time. This will have limited the extent to which the country’s population is represented as such clinics are mostly located in urban centres. However, it avoids the artificial declines in prevalence that could be expected from the expansion of clinics to include those serving more rural populations over recent years. The requirement for settings with consistently sampled clinics will limit the number of places where prevalence trends can be explored. Another way of dealing with problem of sampling biases that change over time is to explicitly explore them. The comparison of random sample population-based surveys with sentinel surveillance data allows the identification of condom use and abstinence.32-33 A similar national level change in patterns of prevalence and incidence was achieved in Thailand. Here, studies that found sex work was the major cause of this behaviour change occurs in year 1990, 1995 or 2000 (arrows). The analyses of HIV trends in Zimbabwe and Kenya were restricted to those antenatal clinics sampled continuously over time. This will have limited the extent to which the country’s population is represented as such clinics are mostly located in urban centres. However, it avoids the artificial declines in prevalence that could be expected from the expansion of clinics to include those serving more rural populations over recent years. The requirement for settings with consistently sampled clinics will limit the number of places where prevalence trends can be explored. Another way of dealing with problem of sampling biases that change over time is to explicitly explore them. The comparison of random sample population-based surveys with sentinel surveillance data allows the identification of condom use and abstinence.32-33 A similar national level change in patterns of prevalence and incidence was achieved in Thailand. Here, studies that found sex work was the major cause of this behaviour change occurs in year 1990, 1995 or 2000 (arrows).
of biases, which can then be included in mathematical models. For example, there is a concern that HIV-associated subfertility will increase over the course of an infection, thereby decreasing the proportion of HIV-infected women becoming pregnant as an epidemic ages. In mathematical models this effect is counteracted by the decreasing bias caused subfertility due to bacterial STIs, which becomes less important after HIV has spread beyond those most likely to have STIs and as HIV-related mortality selectively removes such individuals from the population.\(^{31, 34}\)

Once we are confident that declines in HIV prevalence exceed those expected from the natural dynamics the obvious question is what has caused the declines. Questions of attribution are important as we need to know which risks are amenable to change and how they might be changed. There are two causal levels of interest: first, what change in the distribution of risk explains the reduction in HIV incidence, and second, can these changes in risk be attributed to particular interventions?\(^{35}\) Unfortunately, in Uganda there has been much scope for speculation and appropriation of the success story to support a priori beliefs because of the lack of representative, contemporaneous and detailed risk-behavioural data. However, despite the lack of behavioural data it is possible to draw some conclusions from the observed pattern of prevalence decline. Fitting a model to the decline in Kampala shows that the reduction in aggregate risk must have begun during the late 1980s (fig 2), which indicates it probably was not driven by the increased use of socially marketed condoms that happened later.\(^{32, 33}\) On the other hand, it seems that the dramatic turnaround would not have been possible without all three types of behaviour change across the population—that is, delaying commencing sexual activity by two years, lowering partner change rates and increasing condom use\(^{25, 26, 32}\) (fig 2).

In Zimbabwe the situation is clearer because the observed national-level decline in HIV prevalence\(^{37}\) coincided with population-based surveys of sexual behaviour in the rural eastern highlands\(^{29}\) and observed reductions in incidence among postnatal women and male factory workers in the urban population.\(^{38}\) The former suggested that in these rural communities declines in numbers of sexual partners and an increase in age of sexual debut were associated with the changing incidence, but the latter observed that an increase in condom use was associated with changes in risk in Harare. Behavioural data can be used to identify the components leading to HIV declines when the observed behaviour changes can have a direct influence on the risk of acquisition and transmission of HIV and other similar risk behaviours have not changed. Only if the proximate determinants of HIV infection change can interventions alter the course of the HIV epidemic. The proximate determinants of HIV incidence are those factors influencing\(^{39}:\)

- exposure of uninfected individuals to HIV-positive individuals;
- transmission probability of the virus when such exposure occurs; and
- duration of infectiousness of HIV-positive individuals.

Social, demographic, economic, and other factors that affect HIV incidence ultimately act through their effect on these proximate determinants.

Beyond attributing reduced HIV incidence to particular behaviour changes, relating those behaviour changes to particular interventions requires data measuring exposure to the interventions, perceived responses to those interventions, and plausible mechanisms linking the interventions and the observed changes. Such attribution will always be weak if there are multiple interventions and societal changes. A failure to change risk over the long term may be indicative of unsuccessful interventions, but behaviour change that is consistent with an intervention’s goals suggests that the intervention has worked and that the replication and expansion of the approaches taken might be warranted. Too often such evidence is ignored because it does not meet the requirements of “rigorous” evaluation. Although observed trends in risks within populations do not provide definitive evidence of the failure or success of an intervention generally, the same is true for CRCTs where results from trials of interventions designed to target specific groups or behaviours cannot be generalised.

Use of national-level surveillance data is not suitable for all epidemiological questions. It would not be possible to resolve the difference between the natural, intrinsic course of the epidemic from an epidemic where extrinsic changes alter the course of a growing epidemic (ie when incidence has not already reached a peak) because patterns of risk cannot be adequately assessed in order to predict the peak prevalence. If behaviour changes before the incidence saturates, then the epidemic will settle to a lower level than it would have otherwise but the decline in prevalence will look like the natural dynamics only, leading to the erroneous conclusion that behaviour has not changed (fig 1).

Another problem is the ability of surveys of risk behaviour to capture the details of sexual behaviour necessary to accurately assess the potential for epidemic spread. The distribution of numbers of sexual partners, the pattern of sexual partner choice, the overlap between sexual partners, and the number and type of sex acts within partnerships, all influence the acquisition and transmission of HIV and many advances in study design have improved our ability to measure such detail. However, some uncertainty is inevitable. Details about the sexual contacts of one’s sexual contacts cannot be measured in ordinary population-based surveys. In addition, as the highest-risk individuals are typically a small proportion of the population—and may be under-represented in household surveys (eg sex workers, injecting drug users)—it is difficult to obtain reliable estimates of their behaviour although they contribute disproportionately to the transmission. This matters if declines in HIV incidence are not explained in surveys or if observed changes are insufficient to explain the magnitude of declines. However, it is unlikely that less-obvious changes in

Figure 2. The epidemic in Kampala, Uganda, and the effect of ‘‘ABC’’ interventions. HIV epidemic model fitted (line “ABC”) to antenatal clinic HIV prevalence estimates (black bars with 95% CI). (Source: US Census Bureau.\(^{29}\)) Also shown are epidemic projections assuming no behaviour change (—line), increases in condom use (line “C”), and increase in condom use and two-year average delay in first sex (line “AC”), parameterised according to observed behaviour change.\(^{25}\) The horizontal dashed line shows peak prevalence and is for ease of reading.
behaviour will explain dramatic decreases in HIV incidence, especially as surveys are well designed, with reasonably large sample sizes.

The expansion in funding for HIV prevention activities associated with the creation of the Global Fund for AIDS, Tuberculosis and Malaria** and the (US) President’s Emergency Plan for AIDS Relief (PEPFAR)*** and associated increases in representative HIV surveillance and behaviour data from sentinel sites and from Demographic and Health Surveys with HIV serological surveys (DHS+) has increased both the imperative and the opportunities for testing the large-scale success of interventions. Large-scale interventions must be associated with the collection of sufficiently detailed data to allow their evaluation using mathematical models to determine the magnitude of their impact and for its attribution to different components of the intervention. DHS+ and second-generation surveillance data** should be explored to determine whether it is possible to link behaviour change to particular interventions. The above-described methods using observational studies in conjunction with mathematical models have the advantage of exploring the actual implementation of interventions at the scale of interest, and, if performed rigorously, provide a set of tools for epidemiology and the development of health policy.

ACKNOWLEDGEMENTS

For funding support, TBH thanks the Wellcome Trust, PJW thanks UNAIDS and GPG thanks the UK Medical Research Council. The authors thank Dr Simon Gregson for thoughtful suggestions which helped improve the manuscript.

AUTHOR CONTRIBUTIONS

All the authors developed and contributed to the ideas in the manuscript. TBH programmed and analysed the model advised by GPG and PJW. All authors contributed to drafting the manuscript.

Authors’ affiliations

Timothy B Hallett, Peter J White, Geoff P Garnett, Department of Infectious Disease Epidemiology, Imperial College London, London, UK

Competing interests: None.

REFERENCES


