Published by Oxford University Press on behalf of the International Epidemiological Association *International Journal of Epidemiology* 2007:**37**;77–87

doi:10.1093/ije/dym232

Measuring effectiveness in community randomized trials of HIV prevention

TB Hallett,[1]* GP Garnett,[1] Z Mupamberiyi[2] and S Gregson[1,2]

Accepted 16 October 2007

Background Complicated HIV transmission dynamics make it unclear how to design and interpret results from community-randomized controlled trials (CRCT) of interventions to prevent infection.

Methods Mathematical modelling was used to investigate the effectiveness of interventions to prevent HIV transmission aimed at high-risk groups and factors related to the chance of recording a statistically significant result.

Results Behaviour change by high-risk groups can substantially reduce HIV incidence in the whole population, although its effect is sensitive to the structure of the sexual network and the phase of the epidemic. There is a delay between the behaviour change happening and its full effect being realized in the low-risk group and this can pull the measured incidence rate ratio towards one and reduce the chance of recording a statistically significant result in a CRCT. Our simulations suggest that only with unrealistically favourable study conditions would a statistically significant result be likely with 5 years follow-up or less. Small differences in the epidemiological parameters between communities can lead to misleading incidence rate ratios. Behaviour change independent of the intervention can increase the epidemiological impact of the intervention and the chance of recording a statistically significant result.

Conclusions HIV prevention interventions, especially those targeted at high-risk groups may take longer to work at the population level and need more follow-up time in a CRCT to generate statistically significant results. Mathematical modelling can be used in the design and analysis of CRCTs to understand how the impact of the intervention could develop and the implications this has for statistical power.

Keywords HIV, randomized controlled trials, statistical power

Introduction

Prevention of HIV infection is essential in reducing the morbidity and mortality associated with AIDS and changing the patterns of sexual behaviour has been a major aim of HIV interventions.[1] Whilst a few individual-randomized controlled trials have recorded changes in behaviour,[2,3] demonstrations of successful interventions in reducing HIV incidence at the

[1] Department of Infectious Disease Epidemiology, Imperial College London, UK.

[2] Biomedical Research and Training Institute, Harare, Zimbabwe.

* Corresponding author. Department of Infectious Disease Epidemiology, Imperial College London, St Mary's Campus, Norfolk Place, London, W2 1PG, UK. E-mail: timothy.hallett@imperial.ac.uk

population level are rare. Unfortunately, because the translation of intervention activities into behavioural outcomes and then into a population-level impact on transmission is not inevitable, the key measure of intervention success must be HIV incidence. Some success in reversing the HIV epidemic at a national level was detected in Uganda and Thailand, through an evaluation of trends in prevalence and risk behaviours.[4–7] More recently, similar national trends have been observed in Zimbabwe,[8–10] urban Haiti,[11] parts of urban Kenya[12,13] and south India.[14] In these circumstances the cause of behaviour change is uncertain. To justify the ongoing commitment and support for interventions, it is helpful if their effectiveness has been rigorously demonstrated,[15] and the community-randomized controlled trial (CRCT) is the experimental design that provides a test of an intervention's effectiveness within the population.[16]

There have been four CRCTs to determine the effect of HIV-prevention interventions on HIV incidence, which have run for between two and four years and studied five or six communities per arm, each comprising about 1000 individuals.[17–20] They have all focused on reducing transmission from and to those at most risk of infection (i.e. the 'high-risk' groups—either implicitly through controlling sexually transmitted infections (STIs), which are associated with high risk, or explicitly focusing on sex workers and their clients) but have employed different strategies. In the Mwanza Trial in north-west Tanzania, syndromic management (treating those with symptoms) of STIs was found to reduce HIV incidence by 40%.[17] The Rakai Trial in Uganda attempted to improve on the control of STIs using mass antibiotic administration to target asymptomatic infections[18] and the Masaka Trial, also in Uganda, added a general population Information, Education and Communication (IEC) component to syndromic management of STIs.[19] Most recently, the Manicaland Trial focused on peer-led education and awareness targeted to sex workers and their clients in a rural Zimbabwean community.[20] Unfortunately, none of the latter three (Rakai, Masaka and Manicaland) CRCTs found significant reductions in HIV incidence and in the Manicaland trial incidence may even have been higher in the intervention communities. Explanations as to why these trial results differ from the original success in Mwanza include the epidemiological context, with HIV having spread beyond the high-risk group, in which bacterial STIs play a critical role in transmission.[21,22]

There are many problems in assessing how well complex, multifaceted interventions have been implemented, in estimating how long they need to take effect and in discerning whether the changing epidemiology of HIV has influenced the results. In planning and interpreting trials of HIV prevention interventions, it is worth exploring the expected impact of behavioural change on HIV incidence over time to identify the circumstances when the effectiveness of an intervention would be apparent. Standard power calculations for CRCT[23,24] are available and elaborations have been suggested to optimize study design and analysis.[25–28] However, implications for study design and analysis of the dynamic response of incidence to a targeted intervention has not been considered. To examine this process, we have constructed a simulation of the heterosexual spread of HIV and the impact of a theoretical intervention and compared the chance of detecting a significant effect with estimates from standard power calculations. Whilst our theoretical exploration provides general insights for community-randomized HIV intervention trials, in some cases we parameterize the models to represent aspects of the most recent trial (in Manicaland) in order to ground our analyses in an actual scenario.

Methods

Epidemic simulation model

A mathematical model, parameterized according to observed behaviour in rural Zimbabwe,[9,29] was developed and analysed. The model, based on others,[30] was a deterministic representation of the heterosexual transmission of HIV in a sex, age and sexual activity stratified population. HIV is assumed to be transmitted only at sexual intercourse. The population is divided into three sexual activity groups, representing female sex workers or their male clients (high-risk), those who form casual partnerships and those in stable partnerships (low-risk). On sexual debut, most individuals enter stable partnerships, but a minority enter the higher risk groups. An individual of a particular sex, age and sexual activity forms a set number of sexual partnerships each year. Partnerships are preferentially directed between older men and younger women and between members of the equivalent sexual activity group in the opposite sex.

During the follow-up period, risk behaviour is assumed to decrease independently of the intervention and irrespective of gender, age and sexual activity. To represent this, between year 19 and 24 of the epidemic (representing years 1997–2002), the mean number of sexual partnerships formed per year was reduced by 25% (from starting values between 0.7 and 4.2 depending on age) and the proportion of sex acts in which condoms are used was increased by 2.8 times (from starting values between 10% and 40% depending on age and the type of partnership), based on the recently recorded behavioural changes in Zimbabwe.[9,10] (All analyses were repeated assuming that risk behaviour did not change independently of the intervention to ensure that this did not qualitatively influence the results.)

Against this epidemiological background, we explored how rapidly and clearly the effectiveness of high-risk targeted intervention could be recorded. We assumed that the intervention only directly affected

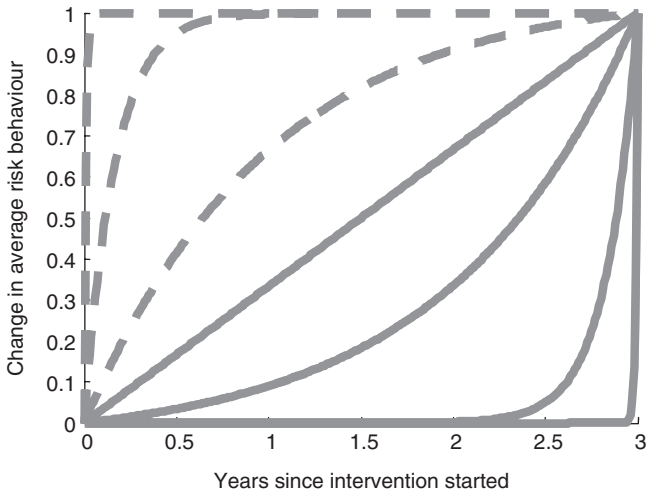

Figure 1 Increasing behaviour change intervention over time. The fraction of the eventual behaviour change completed at time t is $f(t) : f(t) = (e^{v\tau} - 1)/(e^{vT} - 1)$ for $\tau \leqslant T$: τ is the time since the intervention started, T is the duration of time it takes the intervention to reach scale and v is the rate at which the intervention changes behaviour. The intervention can lead to behaviour change occurring quickly ($v < 0$; dashed lines) or more slowly ($v > 0$; solid line); ($v \neq 0$). In these simulations, $T = 3$ and, from top-most line down, the values of v used are -10, -3, -1, 0.1, 1, 3 and 10

female sex workers and their clients (the high-risk group) and generated a 4-fold increase in the frequency with which they used condoms with any partner and reduced by 10% the chance that they had an STI which enhances HIV transmission. These changes are completed within 3 years and the speed of behaviour change can increase or decrease over this period (Figure 1).

Power calculations

Using another model we estimated the chance that the predicted effect of the intervention could generate a statistically significant result after a certain period of follow-up. We approached the problem by running a batch of 5000 stochastic simulations of the intervention trial with the change in incidence output from the dynamic simulation model. At the end of each simulation we assess whether the data that would have been collected would be sufficient to give a statistically significant result (for a two-tailed test at the 5% significance level). Because incidence and losses to follow-up are experienced probabilistically, different trial outcomes are recorded each time, despite assuming the same underlying deterministic effect of the intervention. The proportion of simulations in which a significant result was obtained could be used to estimate the probability of detecting a statistically significant result in one trial (the power of the study design). The hypothesis test is based on the statistic T (equation 1) being distributed under the null hypothesis as t with $n-1$ degrees of freedom.[31]

$$T = \frac{\sqrt{n}\bar{d}}{s_d}$$

$$d_i = \sum_{i=1}^{n} (\log r_{i,2} - \log r_{i,1})$$

$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n}$$

$$s_d = \sqrt{\frac{\sum_{i=1}^{n} (d_i - \bar{d})^2}{n-1}}$$

Both models are fully described in the supplementary material.

Equation 1: T is the test statistic, distributed as t under the null hypothesis with $n-1$ degrees of freedom; n is the number of pairs of communities; $r_{i,j,F}$ is the measured incidence rate in the i-th community in the j-th arm of the study (1 for control communities, 2 for intervention communities) at the end of the follow-up period.

Results

The theoretical effectiveness of an intervention targeted to high-risk individuals

Behaviour change by the high-risk group of women and men (representing sex workers and their clients) can have a substantial impact on HIV incidence in the whole population. With the dramatic behaviour change assumed (quadrupling condom use frequency and 10% reduction in prevalence of co-factor STIs), an intervention started 20 years into the epidemic immediately generates ~70% reduction in incidence in the high-risk group, and incidence in the whole population is reduced by ~35% after 10 years.

An intervention targeted at the high-risk group is likely to have a bigger impact when implemented earlier in the epidemic than when implemented later, because at the beginning of the epidemic most new infections occur in the high-risk group (Figure 2a). The impact of behaviour change by the high-risk group is also sensitive to the structure of the sexual network. Behaviour change by the high-risk group will only be translated into a substantial reduction in incidence in the low-risk group if it is common for high-risk individuals to form sexual partnerships with low-risk individuals ('random mixing') (Figure 2b). If high-risk individuals rarely form partnerships with low-risk individuals ('assortative mixing') then any reduction in incidence in the high-risk group is not shared with the low-risk group.

The reduction in incidence in the whole population is not instantaneous—first, it is assumed that it takes time for average behaviour to change in the targeted group, and second, there is a delay before the reduced

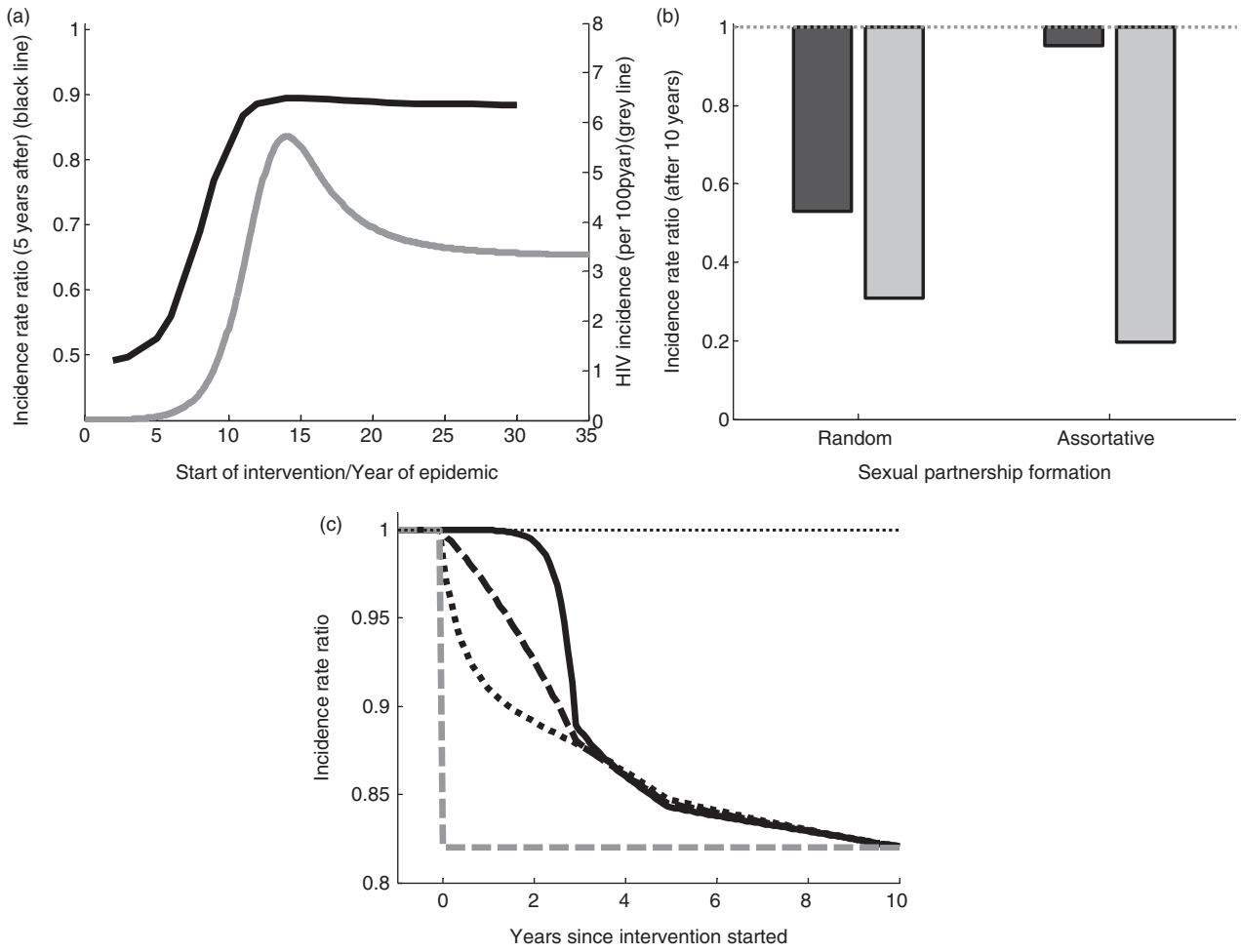

Figure 2 Theoretical effectiveness of an intervention targeted to high-risk individuals (**a**) Predicted IRR 5 years after high-risk targeted intervention starts as a function of the calendar year in which the intervention starts (black line on left-hand vertical axis) and modelled incidence between 1980 and 2015 (grey line on right-hand vertical axis) showing phase of epidemic. (**b**) Predicted IRR among low-risk (dark bars) and the high-risk (light bars) groups 10 years after the high-risk targeted intervention starts if the pattern of sexual partnership formation is 'random' or 'assortative' (see text for definitions). (**c**) IRR as a function of time since high-risk targeted intervention starts if behaviour change is instantaneous (black dotted line), takes 3 years and progresses at a constant rate ($v = 0.01$; black dashed line) or takes 3 years and the rate of change increases ($v = 3$; solid black line). Shown as a comparison, the assumption about IRR made in standard power calculations: IRR is reduced instantaneously (grey dashed line). The pattern of mixing is mid-way between assortative and random (see supplementary material for details)

risk behaviour is translated into a maximum reduction in incidence in the whole population (Figure 2c). The impact of the interventions that reduce transmission from members of the high-risk group drive a faster reduction in incidence in the low-risk group than equivalent interventions that reduce the chance of acquisition (Figure S2). With targeted interventions, the full impact of the intervention is not realized for many years after the intervention starts, even if behaviour changes immediately (Figure 2c—black dotted line). On the other hand, with interventions that affect all members of the community equally, the most important delay will be the time taken for substantial behavioural changes to take hold (Figure S3).

The chance of detecting a statistically significant result for an effective intervention

In practice, the effect of a CRCT is measured as the total number of events during follow-up divided by (estimated) person-years of observation. We will call this the 'measured incidence ratio ratio' (MIRR). However, the potential usefulness of the intervention in programmes is indicated by the *instantaneous* incidence rate ratio (IRR) after the intervention has taken effect (Instantaneous incidence is the rate of new infections per person-year at risk). When rates of infection are low (as for HIV), MIRR will approximate the average of IRR over the follow-up period.

If the full effect of the intervention is exerted immediately, then MIRR would equal IRR. If the full

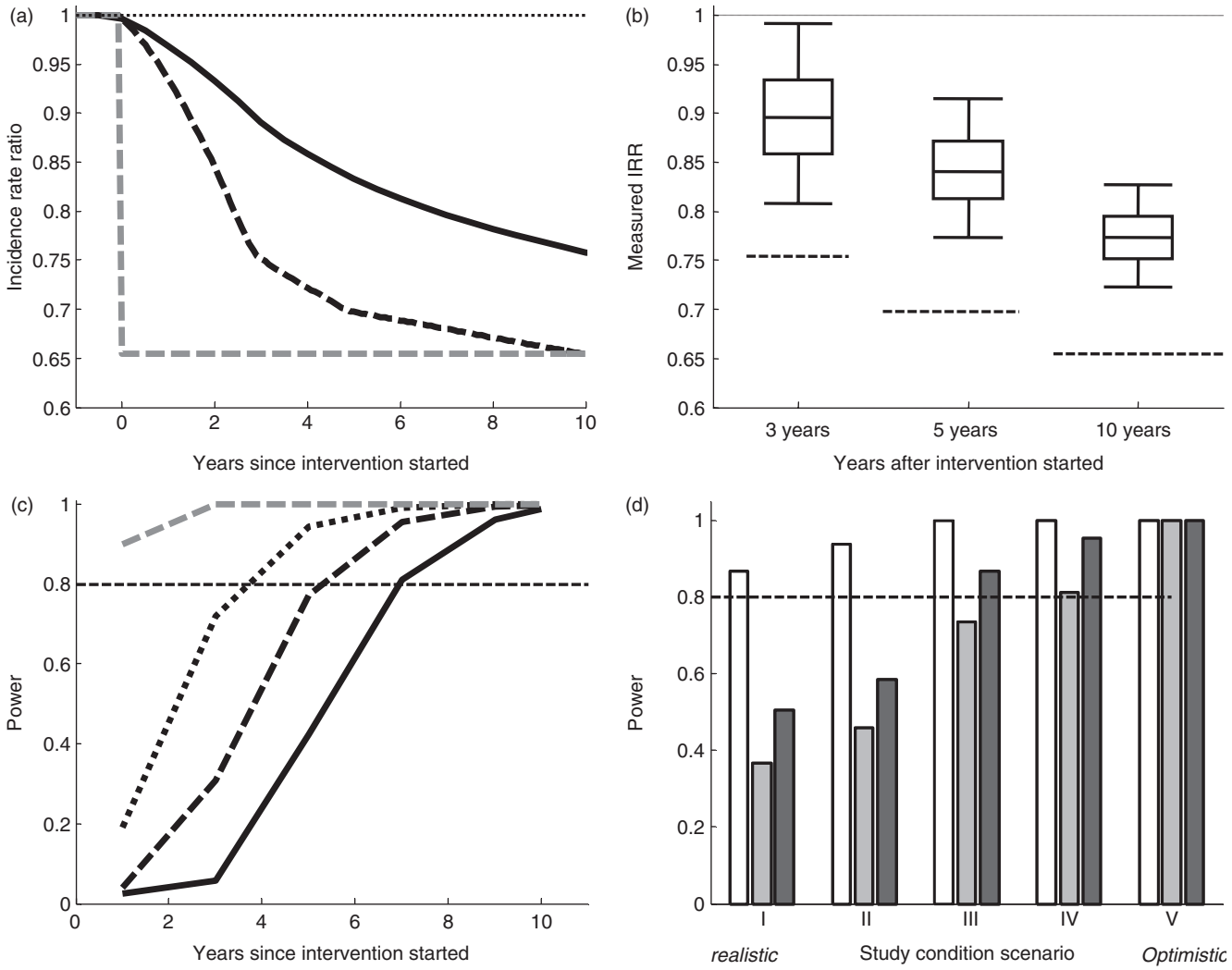

Figure 3 The chance of detecting a statistically significant result for an effective intervention. (**a**) MIRR (black solid line) and IRR (black dashed line) as a function of time since high-risk targeted intervention starts (with constant rate of behaviour change). Shown as a comparison, assumption about IRR made in standard power calculations: IRR is reduced instantaneously (grey dashed line). (**b**) Distribution over 5000 stochastic runs of MIRR after 3, 5 and 10 years follow-up, assuming no between-community variation and no losses to follow-up. Box shows inter-quartile range, horizontal tick indicates median and whiskers extend to 5th and 95th percentiles. Dashed horizontal lines show corresponding IRR. (**c**) Estimated power of the CRCT (two-tailed test at the 5% significance level) as a function of years of follow-up, if behaviour change is instantaneous (black dotted line), takes 3 years and progresses at a constant rate (black dashed line) or takes 3 years and the rate of change increases (solid black line) or if IRR is reduced instantaneously (grey dashed line). Other assumptions are as in (a) and (c). (**d**) Estimated power of the CRCT for a range of study condition scenarios. White bars show estimate based on immediate change in incidence with 5 years follow-up; light grey bars show estimate based on gradual changes in incidence over time with 5 years follow-up; dark grey bars show estimate based on gradual changes in incidence over time with 10 years follow-up. Study condition scenarios are: (I) six communities per arm, 1000 individuals per community, CV of between-community variation = 0.15, 47% followed-up after 5 years; (II) as for I but with 3000 individuals per community; (III) as for II but with 12 communities per arm; (IV) as for III but with 100% follow-up (V) as for IV but with CV = 0.05

effect of the intervention takes time to develop, MIRR will be closer to one than IRR, reflecting a smaller effect of the intervention initially (Figure 3a and b). The size of the discrepancy between the two measures is greatest when the reduction in incidence increases over time. Although in the long-term the two measures converge, over the first years of the intervention, the difference between MIRR and IRR increases as the impact of the intervention grows rapidly (Figure S4).

These three factors (time for behaviour change, time for effect to spread from the high-risk group and measurement of MIRR not IRR) may not be considered in estimates of study power that implicitly

assume an instantaneous reduction in incidence (Figure 3a—dashed grey line). This simplifying assumption will lead to over-estimating the chance that the study returns a statistically significant result for an intervention that works (the power of the study) (Figure 3c). For example, the standard estimate of power for a theoretical 5-year study with ideal study conditions (no loss to follow-up; six identical communities per arm with 1000 individuals per community) could exceed 0.99, but, when we allow for the gradual reduction in incidence, we estimate power to be approximately 0.71. Under more likely study conditions[20] (3 years follow-up with 44% loss to follow up; six communities per arm with 788 individuals per community and between-community coefficient of variation 0.14), whilst power would have been estimated to be 0.82 with an immediate reduction in incidence, it reduces to 0.25 if we assume that incidence reduces gradually.

Given the slow speed with which a substantial reduction in MIRR becomes apparent, the requirements on other aspects of the study design become more stringent if sufficient power is to be achieved with a short follow-up period (Figure 3d). Whilst traditional calculations would suggest that a wide range of study conditions would give a good chance of detecting an effect, our simulations suggest that only with very optimistic assumptions about the number of individuals in each community, between-community variation and the follow-up rate, would there be a good chance (probability >0.8) of recording a statistically significant result within 5 or 10 years. Of these factors, between-community variation and length of follow-up have the strongest influence on power. As has been found for other trial conditions,[23] for a given total number of study participants, increasing the number of communities increases power more than increasing the number of people per community (Table 1).

Table 1 Elements of study design and study conditions that could improve study power

	Estimate of Power (±2 standard deviations)	Relative Increase
Years of follow-up		
3 years (r)	0.29 (0.27–0.30)	1.00
5 years	0.38 (0.37–0.39)	1.32
7 years	0.44 (0.42–0.45)	1.51
10 years	0.50 (0.49–0.52)	1.76
Variation between communities		
CV = 0.15 (r)	0.27 (0.25–0.28)	1.00
CV = 0.10	0.38 (0.36–0.39)	1.42
CV = 0.05	0.52 (0.51–0.53)	1.96
CV = 0.02	0.60 (0.59–0.61)	2.26
Follow-up rate (% after 5 years)		
47% (r)	0.27 (0.26–0.29)	1.00
61%	0.28 (0.27–0.30)	1.04
78%	0.31 (0.29–0.32)	1.12
100%	0.31 (0.30–0.33)	1.15
Number of individuals per community		
1000 (r)	0.27 (0.25–0.28)	1.00
1500	0.31 (0.30–0.32)	1.16
2000	0.33 (0.32–0.35)	1.25
3000	0.34 (0.33–0.35)	1.27
Number of communities per arm		
6 (r)	0.27 (0.25–0.28)	1.00
8	0.32 (0.30–0.33)	1.18
10	0.38 (0.37–0.39)	1.42
12	0.45 (0.44–0.47)	1.69

(CV = coefficient of variation). In each univariate analyses, other parameter values used are indicated by (r); these are based on the conditions of the recent trial in Manicaland, Zimbabwe.[20]

The influence of epidemiological context

The epidemiological background against which the intervention is tested can influence the magnitude of the impact of the intervention and the chance of detecting a statistically significant effect. The potential for epidemic spread can be summarized by R_0, the basic reproductive number.[32] The effect of an intervention that reduces R_0 is greater when R_0 is lower,[32] so we expect that background behaviour changes occurring during the follow-up period could enhance the reduction in incidence associated with the trial intervention (Figure 4a). This is different from declines in incidence and effect associated with the progress of the epidemic (Figure 2a), which occurs against a constant underlying contact pattern.[4,13] Despite there being fewer incident infection observed, our simulations indicate that background behaviour change increases the chance of detecting a statistically significant effect (Figure 4b).

Due to the small number of communities that are randomized it is likely that, by chance, HIV prevalence and incidence in the control and intervention communities is not equal. Such an imbalance means that, even before the intervention, IRR would not equal one. Moreover, if the differences reflect variation in the potential for epidemic spread, then the impact of behavioural changes—whether related to the trial or not—would be greater in communities with lower baseline incidence.[32] This means that the impact of an effective intervention would be underestimated if R_0 is initially higher in the intervention communities, and the impact of an ineffective intervention would be over-estimated if R_0 is initially lower in the intervention communities, especially if there are background behaviour changes. In the Manicaland trial, for instance, mean baseline HIV

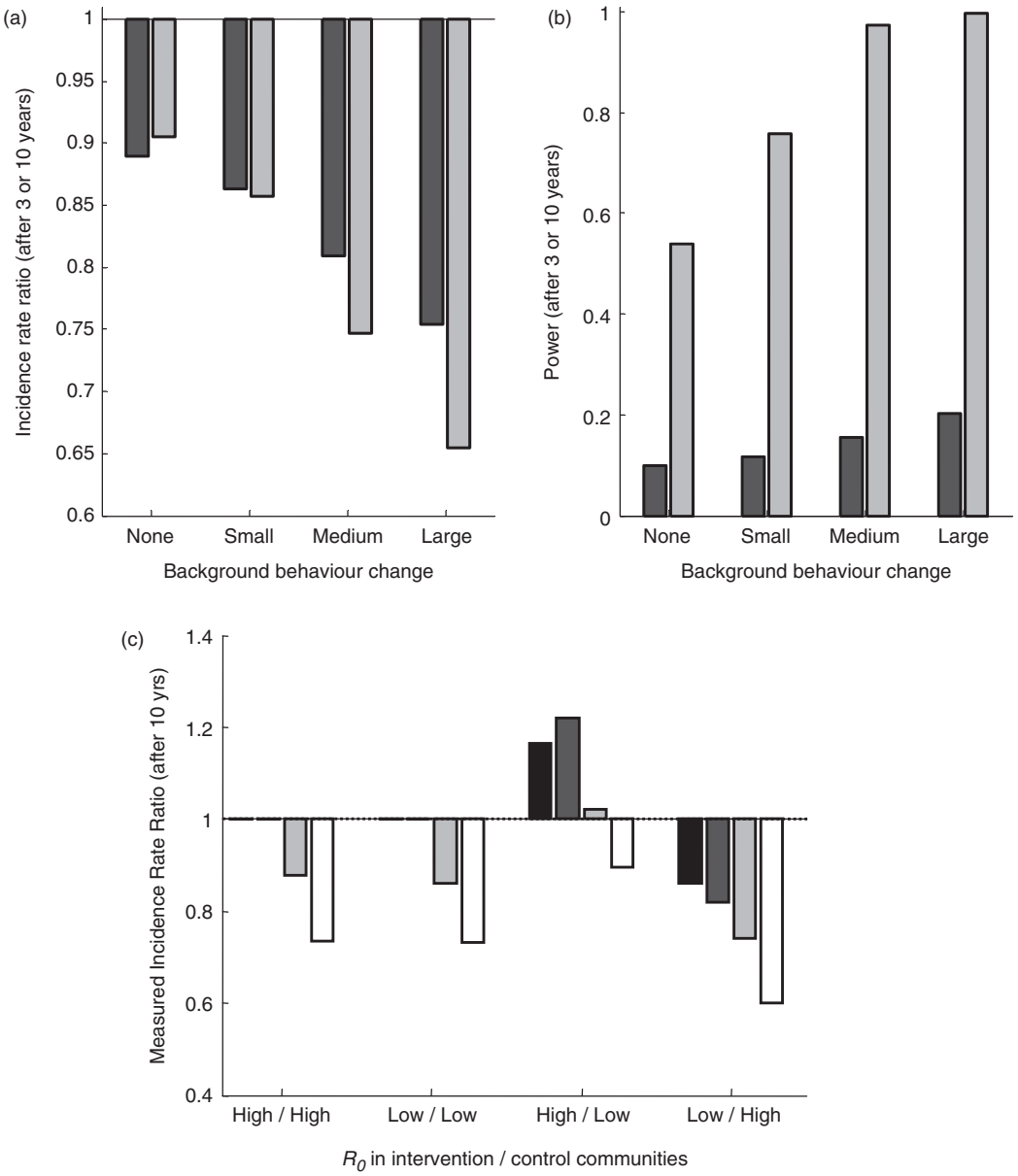

Figure 4 The influence of epidemiological context. (**a**) Predicted IRR of a high-risk targeted intervention after 3 (dark grey bars) or 10 (light grey bars) years assuming that there is no background behaviour change or that condom use increases by 30%, 100% or 280% and partner acquisition rate decrease by 5%, 15% or 25% ('small', 'medium' and 'large' amounts of behaviour change, respectively) linearly throughout the follow-up period equally in all communities. **(b)** Estimated power to detect a statistically significant effect (two-tailed test at the 5% significance level) for the same scenarios as in (a). Study design and conditions assumed to be: six communities per trial arm, 1000 individuals per community, no variation between communities and no losses to follow-up. (**c**) Predicted MIRR after 10 years when there are imbalances in R_0 between intervention and control communities. Assumptions made are: the intervention is not effective and there is no background behaviour change (black bars); the intervention is not effective but there is background behaviour change (dark grey bars); the intervention is effective but there is no background behaviour change (light grey bars); and, the intervention is effective and there is background behaviour change. High and low values of R_0 generate prevalence at the start of the trial of 24% and 21%, respectively

prevalence was 21% in the control communities and 24% in the intervention communities. Simulating this prevalence imbalance by varying R_0 between the control and intervention communities and assuming no background behaviour change, MIRR after 10 years ($MIRR_{10}$) would be 0.86 if R_0 was the same in all communities, but 1.02 if R_0 was higher in intervention communities and 0.74 if R_0 was higher in control communities (Figure 4c). With background behaviour changes, the corresponding $MIRR_{10}$ estimates would be 0.75, 0.90 and 0.60. If the intervention was not effective but there were secular

behavioural changes, $MIRR_{10}$ would be 1.22 if R_0 was higher in intervention communities and 0.82 if R_0 was higher in control communities.

Discussion

The CRCT study design can provide strong evidence of the effectiveness of an intervention, although generalizing the result to all contexts is not necessarily possible.[33] Here to assist in planning and interpreting trials of HIV prevention interventions we have explored the expected impact of behavioural change on HIV incidence over time to identify the circumstances when the effectiveness of an intervention would be apparent. The scientific interpretation of a trial that returns a negative or null result must be informed by these considerations.

Our model suggests that a tendency for random mixing (for example, high-risk men forming sexual partnerships with low-risk young women[29]) would allow a reduction in incidence among the high-risk group to spread throughout the population. However, in a CRCT, this would not necessarily be reflected in a substantial or statistically significant reduction in MIRR after a short follow-up period. The reasons for this are (i) the intervention may take some time to get to scale or to change the behaviour of the high-risk group, and (ii) there is a delay before the reduced risk behaviour by the high-risk group is translated into a reduction in incidence in the whole population. Delays in reductions in incidence will reduce the difference between the measured incidence rates in the two study arms. This means that evidence for the effectiveness of interventions that reduce incidence gradually over a few years—but which could avert many infections in the longer term in actual programmes—might not be found in CRCTs.

One of the purpose of CRCTs is to establish whether an intervention 'works' and should be included in programmes. Estimating the chance that an epidemiologically relevant reduction in incidence would be detected statistically is an integral part of CRCT study design and standard methods for power calculations have been published.[23,24] However, our simulations show that these calculations can over-estimate power if they assume that the reduction in incidence in the whole population is immediate. This conclusion is likely to be conservative since our model of the transmission dynamics is deterministic and does not take account of the chance events that would likely further slow the impact of the intervention. Implicitly taking account of this effect when deciding the predicted effect size in standard power calculations can be misleading because it masks the assumptions (i.e. degree of optimism) about the actual long-term impact of the intervention. Over-estimating the power of a study may encourage the interpretation of a non-significant result as evidence for lack of effect; our simulations suggest that a probable explanation for a non-significant result will often be the slow speed with which the effect of the intervention develops. Mathematical simulation modelling provides a convenient way to examine how the study power is related to study conditions, and to consider the effect of incidence declining gradually over time, whether this is due to the intervention starting more slowly than planned or patterns of transmission changing gradually.

It is difficult to say which of the different ways a trial can improve its ability to detect a significant effect will be the most practical. The choice of communities will probably be determined by many practical matters besides their similarity and there may be a trade-off between using two similar communities close to each other or two that are less similar but further apart, which may reduce contaminating migration. There is also the danger that pairing of communities by attributes that do not correlate well with the true incidence rate could be counter-productive when dealing with a small number of communities.[34,35] Increasing the total number of individuals followed-up in a study will be associated with greater costs, but this will be spent more effectively if the number of communities is increased rather than the number of individuals in each community.[23] Increasing the number of communities also help minimize the danger that, by chance, the two study arms do not have the same basic reproductive number (R_0) and our simulations show that even small differences at baseline can lead to large biases in MIRR. In the extreme cases, this could completely obscure the impact of truly efficacious intervention, or artificially generate a positive result for an ineffective intervention.[36] Generally, statistically controlling for differences in prevalence between sites will be insufficient to eliminate this bias because not only does the value of R_0 determine the expected incidence rates under the null hypothesis, but it can also modify the actual the effect of the intervention and background behaviour changes. This leads to greater variation in the observed effects of the intervention than can be explained by variation in baseline prevalence, which itself will not fully capture variation in R_0.

Losses at follow-up might be reduced by more frequent visits but our simulations suggest that, when taking other difficulties into account, a substantial increase in power is unlikely even with almost no losses to follow-up. Furthermore, if loss to follow-up is mostly accounted for by migration and mortality, as was the cases in Manicaland,[20] then extra visits will do little to help. On the other hand, having an interim round to recruit new migrants could help boost power and allow a measure of incidence later in the study, when the intervention is to scale and beginning to have an effect. Using detuned assays[37] at the follow-up round would also provide a way of measuring incidence later in the follow-up period,

but the statistical uncertainty associated with estimates based on small numbers of recent seroconversions would limit the usefulness of this approach without following-up many more individuals.

Increasing the length of the follow-up period may be the most certain way to increase the chance that the study captures the effect of an intervention. However, commitment to a trial for a decade or longer may be difficult for funding agencies and investigators and runs the risk that the findings of the study will be out-of-date by the time it is complete. These difficulties may mean that, although a sufficiently powered CRCT is the best way to test the effectiveness of an intervention, observational studies based on sentinel surveillance may be increasingly relied upon to gauge the effect of on-going large-scale interventions without control groups.[38]

Some trials have been set against a background of generalized behaviour change that took place equally in all communities and was independent of the intervention. It has been suggested that declining HIV incidence would mask the impact of a successful intervention because there are fewer infection events in the control community, which it is hoped will be prevented in the intervention community.[19,20] Conversely, our results suggest that interventions can have a larger effect when other risk behaviour is reduced (including biomedical changes such as male circumcision) and this can outweigh the difficulties involved in establishing statistical differences between small numbers. Therefore, it is unlikely that secular changes in behaviour obscured the impact of the interventions in the African CRCTs, but this remains a concern in other settings where the incidence of HIV is lower.

The model we have used provides a simple description of the heterosexual transmission of HIV.[30] Key aspects of HIV epidemiology are represented, such as the distribution in risk behaviour and variable infectiousness, but it does explicitly include many other factors such as transmission of STIs or partnership duration and concurrency.[39,40] However, very simple mathematical models can provide important qualitative insights to complicated problems.[41] Although some aspects of the model have been parameterized to reflect the design and epidemiological situation of the most recent CRCT (in Manicaland, Zimbabwe[20]), we believe that our findings will be broadly relevant to other settings and types of study.

CRCTs of HIV prevention interventions are more susceptible to these problems than other infectious diseases, particularly those that focus on those with the riskiest sexual behaviour (e.g. those with STIs or involved in sex work). First, the private and psychologically substantive nature of sexual risk behaviour will tend to make changes slower. Next, the distribution of risk and pattern of contact with respect to risk may be more likely to be assortative (delaying the spread of the effect) than other forms of contact.[42] Third, the long duration of HIV infection delays the impact of intervention spreading beyond the highest-risk group, particularly if the changes reduce the chance of infection acquisition rather than transmission (e.g. treating ulcerative STIs among women,[43] or male circumcision). If the chance of acquisition is reduced and the infection lasted less time, prevalence among the high-risk group would decline faster, and the risk of contact with that group would reduce more quickly. If the chance of transmission is reduced, the speed with which incidence is reduced in the population is most limited by the extent of contact between high and low-risk individuals and not the duration of infection (Figure S1). Added to this is the fact that incident HIV infection is a relatively rare outcome in the general population and the chance of infection in intervention and control sites is subject to chance, epidemic phase, secular trends driven by natural dynamics[13] and/or widespread behaviour change, all of which may differ slightly between superficially similar communities.

Other infectious diseases may share one or more of these attributes and therefore face a similar problem. For example, sexual behaviour change to reduce the incidence of other STIs may also take time to work and the impact of interventions to prevent or treat tuberculosis will be slow to develop because there is a slow turnover of the infected population. On the other hand, interventions that aim to reduce the population of insect vectors of diseases such as malaria may have fewer problems since the turn-over of vectors is high and an effective reduction in the risk of infection to humans can develop quickly.[32]

The CRCT is the gold standard for demonstrating the effectiveness of an intervention in a specific context.[15,16] If a negative or null result is obtained, after assessing how well the intervention was realized, one has to consider whether the intervention was not appropriate for the epidemiological context or whether the study design (as implemented) was able to detect a real effect of epidemiological importance. Mathematical descriptions of the transmission of HIV can help explore how efficacious the intervention could be, but to rule out the chance the study failed to detect an effect, CRCTs for the prevention of HIV infection in generalized epidemics may need to include more communities or run for much longer.

Supplementary data

Supplementary data are available at *IJE* online.

Acknowledgements

T.B.H., Z.M. and S.G. thank The Wellcome Trust and G.P.G. thanks The Medical Research Council for funding support. The authors are grateful to N Grassley and P White for useful discussion.

Conflict of interest: None declared.

KEY MESSAGES

- The impact of interventions to prevent HIV infections can take time to develop, especially if they mostly operate through those with the riskiest sexual behaviour. Trials need to allow time for this.
- Targeted interventions have a greater impact early in an epidemic and when mixing between the target group and the rest of the population is extensive.
- Independent reductions in risk across trial communities can increase rather than decrease the power of a community randomized controlled trial.

References

[1] UNAIDS. 2006. *Report On The Global Aids Epidemic*. Geneva: UNAIDS, 2006.

[2] The Voluntary HIV-1 Counseling and Testing Efficacy Study Group. Efficacy of voluntary HIV-1 counselling and testing in individuals and couples in Kenya, Tanzania, and Trinidad: a randomised trial. *Lancet* 2000;**356:**103–12.

[3] Kamb ML, Fishbein M, Douglas JM Jr *et al*. Efficacy of risk-reduction counseling to prevent human immunodeficiency virus and sexually transmitted diseases: a randomized controlled trial. Project RESPECT Study Group. *JAMA* 1998;**280:**1161–67.

[4] Kilian AH, Gregson S, Ndyanabangi B *et al*. Reductions in risk behaviour provide the most consistent explanation for declining HIV-1 prevalence in Uganda. *Aids* 1999;**13:**391–98.

[5] Stoneburner RL, Low-Beer D. Population-level HIV declines and behavioral risk avoidance in Uganda. *Science* 2004;**304:**714–18.

[6] Nelson KE, Celentano DD, Eiumtrakol S *et al*. Changes in sexual behavior and a decline in HIV infection among young men in Thailand. *N Engl J Med* 1996;**335:**297–303.

[7] Rojanapithayakorn W, Hanenberg R. The 100% condom program in Thailand. *Aids* 1996;**10:**1–7.

[8] Mahomva A, Greby S, Dube S *et al*. HIV prevalence and trends from data in Zimbabwe, 1997-2004. Sex Transm Infect 2006;**82 (Suppl 1):**i42–47.

[9] UNAIDS. Evidence for HIV decline in Zimbabwe: a comprehensive review of the epidemiological data; 2005.

[10] Gregson S, Garnett GP, Nyamukapa CA *et al*. HIV decline associated with behavior change in eastern Zimbabwe. *Science* 2006;**311:**664–66.

[11] Gaillard EM, Boulos LM, Andre Cayemittes MP *et al*. Understanding the reasons for decline of HIV prevalence in Haiti. *Sex Transm Infect* 2006;**82 (Suppl 1):**i14–20.

[12] Cheluget B, Baltazar G, Orege P, Ibrahim M, Marum LH, Stover J. Evidence for population level declines in adult HIV prevalence in Kenya. *Sex Transm Infect* 2006;**82 (Suppl 1):**i21–26.

[13] Hallett TB, Aberle-Grasse J, Bello G *et al*. Declines in HIV prevalence can be associated with changing sexual behaviour in Uganda, urban Kenya, Zimbabwe, and urban Haiti. *Sex Transm Infect* 2006;**82 (Suppl 1):**i1–i8.

[14] Kumar R, Jha P, Arora P *et al*. Trends in HIV-1 in young adults in south India from 2000 to 2004: a prevalence study. *Lancet* **367:**1164–72.

[15] Stephenson J, Imrie J. Why do we need randomised controlled trials to assess behavioural interventions? *Br Med J* 1998;**316:**611–13.

[16] Susser M. Some principles in study design for preventing HIV transmission: rigor or reality. *Am J Public Health* 1996;**86:**1713–16.

[17] Grosskurth H, Mosha F, Todd J *et al*. Impact of improved treatment of sexually transmitted diseases on HIV infection in rural Tanzania: randomised controlled trial. *Lancet* 1995;**346:**530–36.

[18] Wawer MJ, Sewankambo NK, Serwadda D *et al*. Control of sexually transmitted diseases for AIDS prevention in Uganda: a randomised community trial. Rakai Project Study Group. *Lancet* 1999;**353:**525–35.

[19] Kamali A, Quigley M, Nakiyingi J *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:**645–52.

[20] Gregson S, Adamson S, Papaya S *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.

[21] White RG, Orroth KK, Korenromp EL *et al*. Can population differences explain the contrasting results of the Mwanza, Rakai, and Masaka HIV/sexually transmitted disease intervention trials?: a modeling study. *J Acquir Immune Defic Syndr* 2004;**37:**1500–13.

[22] Orroth KK, Korenromp EL, White RG *et al*. Higher risk behaviour and rates of sexually transmitted diseases in Mwanza compared to Uganda may help explain HIV prevention trial outcomes. *Aids* 2003;**17:**2653–60.

[23] Donner A, Klar N. *Design and Analysis of Cluster Randomization Trials in Health Research*. London: Hodder Arnold Publication, 2000.

[24] Hayes RJ, Bennett S. Simple sample size calculation for cluster-randomized trials. *Int J Epidemiol* 1999;**28:**319–26.

[25] Localio AR, Berlin JA, Have TR. Longitudinal and repeated cross-sectional cluster-randomization designs using mixed effects regression for binary outcomes: bias and coverage of frequentist and Bayesian methods. *Stat Med* 2006;**25:**2720–36.

[26] Moerbeek M. Power and money in cluster randomized trials: when is it worth measuring a covariate? *Stat Med* 2006;**25:**2607–17.

[27] Eldridge SM, Ashby D, Kerry S. Sample size for cluster randomized trials: effect of coefficient of variation of cluster size and analysis method. *Int J Epidemiol* 2006;**35:**1292–300.

[28] Campbell MJ, Donner A, Klar N. Developments in cluster randomized trials and statistics in medicine. *Stat Med* 2007;**26:**2–19.

[29] Gregson S, Nyamukapa CA, Garnett GP *et al*. Sexual mixing patterns and sex-differentials in teenage exposure to HIV infection in rural Zimbabwe. *Lancet* 2002;**359:**1896–903.

[30] Garnett GP, Anderson RM. Factors controlling the spread of HIV in heterosexual communities in developing countries: patterns of mixing between different age and sexual activity classes. *Philos Trans R Soc Lond B Biol Sci* 1993;**342:**137–59.

[31] Brookmeyer R, Chen YQ. Person-time analysis of paired community intervention trials when the number of communities is small. *Stat Med* 1998;**17:**2121–32.

[32] Anderson RM, May RM. *Infectious Diseases Of Humans*. Oxford: Oxford University Press, 1991.

[33] Grassly NC, Garnett GP, Schwartlander B, Gregson S, Anderson RM. The effectiveness of HIV prevention and the epidemiological context. *Bull World Health Organ* 2001;**79:**1121–32.

[34] Martin DC, Diehr P, Perrin EB, Koepsell TD. The effect of matching on the power of randomized community intervention studies. *Stat Med* 1993;**12:**329–38.

[35] Freedman LS, Green SB, Byar DP. Assessing the gain in efficiency due to matching in a community intervention study. *Stat Med* 1990;**9:**943–52.

[36] Korenromp EL, White RG, Orroth KK *et al*. Determinants of the impact of sexually transmitted infection treatment on prevention of HIV infection: a synthesis of evidence from the Mwanza, Rakai, and Masaka intervention trials. *J Infect Dis* 2005;**191 (Suppl 1):**S168–78.

[37] Parekh BS, Kennedy MS, Dobbs 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.

[38] Hallett TB, White PJ, Garnett GP. The appropriate evaluation of HIV prevention interventions: from experiment to full scale implementation. *Sex Transm Infect* 2007;**83:**i55–60.

[39] Kretzschmar M, Dietz K. The effect of pair formation and variable infectivity on the spread of an infection without recovery. *Math Biosci* 1998;**148:**83–113.

[40] Morris M, Kretzschmar M. Concurrent partnerships and the spread of HIV. *Aids* 1997;**11:**641–48.

[41] Aral SO, Roegner R. Mathematical modeling as a tool in STD prevention and control: a decade of progress, a millennium of opportunities. *Sex Transm Dis* 2000;**27:**556–57.

[42] Edmunds WJ, Kafatos G, Wallinga J, Mossong J. Mixing patterns and the spread of close-contact infectious diseases. *Emerg Themes Epidemiol* 2006;**3:**10.

[43] Rottingen JA, Cameron DW, Garnett GP. A systematic review of the epidemiologic interactions between classic sexually transmitted diseases and HIV: how much really is known? *Sex Transm Dis* 2001;**28:**579–97.