ELSEVIER

Social Science & Medicine 63 (2006) 1135–1142

SOCIAL SCIENCE & MEDICINE

www.elsevier.com/locate/socscimed

Should structural interventions be evaluated using RCTs? The case of HIV prevention

Christopher Bonell*, James Hargreaves, Vicki Strange, Paul Pronyk, John Porter

Department of Public Health and Policy, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK

Available online 12 May 2006

Abstract

Structural interventions addressing macro-social determinants of risk have been suggested as potentially important adjuncts to biomedical and behavioural interventions for the prevention of HIV and other diseases. A few interventions of this type have been evaluated using randomised controlled trials (RCTs), the most rigorous design to evaluate the effects of biomedical and behavioural interventions. The appropriateness of applying RCTs to structural interventions is however debated.

This paper considers whether issues of ethics, feasibility and utility preclude the use of RCTs in evaluations of structural interventions for HIV prevention. We conclude there is nothing particular to this category of interventions prohibiting use of RCTs. However, we suggest that RCTs may prove unacceptable, unfeasible or not useful in certain circumstances, such as where an intervention brings important benefits other than HIV prevention (such as increased income); where leaders of clusters do not allow decisions about macro-social policies to be determined randomly; where the unit of social organization addressed by an intervention is so large that recruitment of adequate numbers of clusters is impossible; and where the period required to trial interventions extends beyond practical decision-making time-scales. In such cases, alternative evaluative designs must be assessed for their ability to provide evidence of intervention effectiveness.

Keywords: Structural; Intervention; HIV; Prevention; Evaluation; Randomised control trials

Structural factors influencing HIV transmission and interventions to address these

Approximately 38 million people were living with HIV in 2004 (UNAIDS, 2004). Research has identified factors influencing HIV transmission at the biomedical and behavioural levels but also at the structural level. Structural interventions have been suggested as a potentially important adjunct to biomedical and behavioural interventions for the prevention of HIV and several have been deployed (Farmer, Connors, & Simmons, 1996; RADAR, 2002a, b; Tawil, Verster, & O'Reilly, 1995; Zierler & Krieger, 1997).

This paper explores whether the effects of structural interventions can and should be evaluated using randomized controlled trials (RCTs), using HIV prevention to focus discussion. We first define and give examples of structural interventions. We then consider the RCT design itself and what features are required for RCTs of structural interventions, before discussing the ethics, feasibility and utility of conducting such trials. Finally we

*Corresponding author. Tel.: +44 20 7927 2415.
E-mail address: chris.bonell@lshtm.ac.uk (C. Bonell).

0277-9536/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.socscimed.2006.03.026

briefly consider alternatives to RCTs where the latter are not ethical, feasible or useful.

Structural interventions

The 'broad-school' definition (Rhodes, Singer, Bourgois, Friedman, & Strathdee, 2005: S11) of structural interventions encompasses those "that work by altering the context in which health is produced or reproduced … [and which] locate the source of public-health problems in factors in the social, economic and political environments that shape and constrain individual, community, and societal health outcomes". Structural interventions are thus those that address factors outside individual control (Sumartojo, 2000). Broad-school definitions encompass interventions aiming to foster 'macro-social' changes (e.g. in the distribution of power and resources, the running of major institutions and legal systems and public policies) as well as those addressing 'meso-social' change (e.g. social networks, community norms and local health systems). The narrow-school definition (Rhodes et al., 2005) encompasses only interventions addressing macro-social influences. Key macro-social factors influencing HIV transmission are gender inequity, poverty and migration (Parker, Easton, & Klein, 2000). Interventions addressing these might be delivered globally, nationally or locally. Local interventions include ones aiming to reduce social and economic vulnerability to HIV through the provision of micro-finance to households in poverty, delivered at the village-level (RADAR, 2002a, b) or aiming to improve young people's engagement with schools (Patton, Bond, Butler, & Glover, 2003). Such interventions are structural because they address 'upstream' (i.e. causally distal) influences on health (Sorensen, Emmons, Hunt, & Johnston, 1998). Table 1 provides several examples of narrow-school structural HIV prevention interventions referred to in this paper.

The potential relevance of RCTs to evaluate structural interventions

While structural interventions are increasingly regarded as important for HIV prevention, their impact on HIV incidence or risk behaviours has rarely been evaluated (Blakenship, Bray, & Merson, 2000). Some argue that, where feasible, RCTs should be used to evaluate the effects of structural interventions for HIV prevention in order to inform decisions about their deployment (Berkman, 2004). Such trials have indeed been undertaken (Patton et al, 2003; RADAR, 2002a, b). However, others are sceptical about the suitability of the design for evaluating structural interventions (Des Jarlais, 2000; Sumartojo, 2000).

Table 1
Contemporary HIV transmission patterns, structural factors and structural interventions for HIV prevention

Setting/mode of HIV transmission	Example of a structural factor influencing HIV transmission	Examples of potential structural interventions
(1) South Africa/heterosexual transmission among young people	Poverty and gender inequality. Impact mediated for young women by sexual relationships being an important source of potential income and support, and they may be badly placed within these to negotiate safer sex	Change in global aid or trade policies Legislative changes to promote higher incomes and employment for women Provision of basic income grants or increasing access to microfinance credit for poor households
(3) UK/homosexual transmission among men	Legislative discrimination against gay and homosexually active men in areas such as partner rights and employment. Impact mediated by reinforcement of social norms of homophobia influencing self-esteem and sexual negotiation skills of gay and other homosexually active men	Repeal of legislation Policies to address homophobic bullying in schools
(2) USA/use of shared equipment among injecting drug users	Policing policy and practice regarding possession of drug-injecting equipment. Impact mediated by decreased procurement of clean injecting equipment as this risks being found in possession and prosecuted	Work to amend police policy or practice regarding possession of drug-injecting equipment

An RCT is an experiment in which participants undergo random allocation into groups that receive or do not receive an intervention. Later, comparison is made of outcomes between these groups. Within biomedical, behavioural and health services research, the RCT design is considered the most rigorous for assessing intervention effects since it has the potential to minimize bias and confounding that affect other study designs (Moher, Schultz, Altman, & CONSORT Group, 2001).

Although RCTs have long been used to evaluate social interventions, i.e. those operating largely via inter-personal processes, of which structural interventions are a sub-set, this remains controversial (Susser, 1996). Some argue that RCTs are mired in so-called 'positivist' assumptions rendering them unsuitable for evaluating social interventions. However, we suggest that although social phenomena differ from biological ones, it is possible to examine the effects of social interventions using RCTs without being 'positivist'. Our previously made epistemological arguments (Bonell, Bennett, & Oakley, 2003) will not be repeated here. Instead, we focus on the appropriateness of evaluating structural interventions using RCTs.

Whether used to evaluate biomedical, behavioural or structural interventions, RCTs must be rigorously designed, conducted and reported. RCT protocols should specify pre-defined hypotheses, define an a-priori analytic plan and report on all pre-specified outcomes. Consent rates and degree of attrition within arms must be reported. Adherence to the allocated intervention should be recorded. Data collection and randomization methods should be described. RCTs should be overseen by independent data-monitoring committees which assess preliminary data to determine whether trials should be halted either because interventions are clearly effective, so that denying them to the control group is unethical, or because interventions are harmful, rendering it unethical to continue providing them to the intervention group.

RCTs deployed in evaluating structural interventions will generally require a number of specific features as well as those listed above. Firstly, since structural interventions generally operate across whole environments, even if their aims relate to specific population sub-groups, RCTs will usually require cluster-allocation rather than individual-allocation. In such trials, institutions, administrative areas, etc. comprise the clusters and unit of randomization. Individuals within clusters might be affected by the intervention directly, indirectly or both. RCTs need to consider the appropriate groups in whom outcome measurement should be made, these being the target-group among whom health benefits are hypothesised. For example, in example 1 (Table 1), outcomes would be measured among young people residing in the homes or communities of those who receive microfinance.

Secondly, although the primary aim of a trial should be hypothesis-testing or estimating the size of effect of an intervention (Susser, 1996), trials of structural interventions may also aim to build hypotheses through investigation of a wide range of secondary and pathway outcomes in order to understand complex causal chains. This is because structural interventions are likely to work via multiple and extensive pathways. Developing a conceptual framework at an early stage of study design can guide the choice of pathway variables to be measured and help formulate an a-priori analysis plan (Victora, Huttly, Fuchs, & Olinto, 1997). Where such analyses are performed, care must be taken to avoid bias and type-1 error in the analysis and interpretation of data.

Finally, RCTs of structural interventions should include process evaluations. Structural interventions are likely to show variability in local implementation. Process evaluation is important to assess intervention delivery and receipt, and explore how this is related to context (Wight & Obasi, 2003). Combined analysis of process and outcome data should allow evaluations to explore associations between intervention delivery and outcomes, and to report on the likely generalisability of intervention effects.

The features outlined above are not unique to structural intervention trials; some are necessary within RCTs of behavioural and biomedical interventions. For example, vaccines and certain other biomedical interventions exert effects by inter-individual processes and so require cluster-RCTs (Hayes, Alexander, Bennett, & Cousens, 2000). Equally, drug trials may explore pathways of causality (Friedman, Furberg, & DeMets, 1998). Nevertheless, we contend that cluster-RCTs incorporating process evaluations provide the most rigorous design to measure the impact of structural interventions and inform whether these should be deployed. We now consider when in practice it is ethical, feasible and useful conduct to RCTs of structural interventions.

Factors to assess in considering RCTs of structural interventions

Ethical issues

The necessity for evaluating health interventions stems from the ethical imperative to do good and do no harm. Unless the effects of an intervention are immediate and obvious (which is rare) or have been rigorously evaluated, 'equipoise' exists so that evaluation is ethically indicated and use of RCTs is justified.

Most health interventions aim to bring about health benefit without trying to influence significantly other aspects of recipients' lives. A decision whether or not to evaluate such interventions depends on whether there is equipoise concerning health effects. However, structural interventions that address 'upstream' influences on health aim to bring about other benefits, such as enhanced socioeconomic status or civil rights. In some cases, equipoise may exist concerning health, but not other, intended effects. In example 1 (Table 1), it is widely accepted that providing microfinance to poor households can produce economic benefits (Johnson & Rogaly, 1997) but little is known about HIV-effects. In example 2, although the impact of repealing anti-gay laws on vulnerability to HIV infection is uncertain, the civil-rights impact is obvious. Where there is equipoise concerning the health but not other effects of structural interventions, the ethical situation is complex. In some cases, it might be judged that it is not ethically required to evaluate health effects since it is already known that the interventions bring about other important benefits, even if the health effects are uncertain. However, precedents should be borne in mind of interventions assumed to have obvious and multiple benefits but that actually had deleterious effects on health (Oakley, 2000).

RCTs involve denying the intervention under study to the comparison group. We suggest that where there is evidence to suggest that the intervention would bring about non-health benefits to those in the comparison group, it may still be ethical to undertake an RCT. Judgment would depend on the nature of the intervention's other benefits and existing accessibility to the proposed study population. There is consensus among ethicists that humans are entitled to a generalised standard of certain benefits. For example, most ethicists would agree that all humans are entitled to freedom from sexual discrimination (example 2). No such consensus exists regarding access to micro-finance (example 1). Therefore an RCT of the HIV-prevention effects of protecting the civil rights of gay men might be considered unethical, whereas a trial involving micro-finance might be judged ethical. However, it may still be unacceptable, and arguably unethical, to deny an intervention such as provision of micro-finance to study participants if the intervention is already widely accessible.

Thus, researchers and ethics committees must consider: what is already known about the effects of the intervention; whether or not any known intervention benefits are regarded as universal rights; and whether the intervention to be trialled is already accessible to the population being studied. Judgments on each of these issues will not be simple and should be made by researchers and ethicists working together. As we will discuss later, even where ethics committees have no objections to a trial, it may still be unfeasible to conduct one if participants or cluster leaders object.

Where an RCT is to be conducted, a further issue affecting both the ethics and interpretation of the study is the services provided to the comparison group. Trials of HIV prevention interventions require that comparison groups be provided with effective HIV prevention practically accessible as a local standard of care (McQueen & Sugarman, 2003). This should also ensure that trials provide useful evidence, identifying the effectiveness of new interventions compared with the current best option. Again, the situation is more complex with structural interventions that aim to bring about non-health benefits. We suggest that trials of structural interventions that aim to bring other non-health benefits should additionally ensure comparison group-members are not denied services that are currently available in the study site and which are judged likely—according to current evidence—to bring about these other benefits. Fears that have arisen with regard to drug trials that low-income settings will become cheap testing grounds for interventions designed for deployment in high-income settings, and associated arguments about whether comparison groups should receive universal or local standards of care (Angell, 1997), will apply less often to structural or other social intervention trials both because of the absence of for-profit research and the more uncertain transferability of interventions between settings.

Finally, those intending to trial structural interventions will need to consider ethical guidelines for cluster-RCTs. In particular, cluster randomized trials require consideration of which individuals trialists should engage with to discuss participation and what precisely is the role of these 'gatekeepers', it certainly not being analogous or reducible to the giving of informed consent by each participant in individual-allocation RCTs. Consideration is necessary both of individual autonomy and the utilitarian welfare of the whole cluster (Edwards, Braunholtz, Lilford, & Stevens, 1999).

Feasibility

Feasibility refers to the possibility that an RCT design can practically be applied. As discussed earlier, some trials cannot and should not proceed because they are unethical. However, trials can be unfeasible when there are no ethical objections. For example, 'gatekeepers' or participants may simply refuse to consider the possibility of variable provision within a trial regardless of the evidence for the effects of this.

Alternatively, variable provision may be acceptable, but randomization being used to achieve this will not (Des Jarlais, 2000). Cluster-members or gatekeepers may be unwilling to commit themselves to participation in an evaluation before they know whether they will be allocated to the intervention or control arm. This might particularly be a problem when structural intervention trials require those in power within a cluster to obey the dictates of random allocation when it comes to changing the laws, policies and other tools with which they wield power, rather than the provision of micro-level interventions to cluster-members, about which leaders may feel less 'ownership' (Edwards et al., 1999). In example 3, it may prove difficult to undertake a trial of the effects of policing policy on HIV transmission among injecting drug users because this will require local police chiefs' consent, in effect, to relinquishing control over this area, allowing it instead to be randomly determined.

We should stress that the barriers discussed above can often be surmounted. Investigators may be able to persuade communities or institutions of the benefits of participation in such trials if they can convince them that the importance of answering questions about the impact of the intervention, for example on HIV transmission, outweighs the personal opinions and ideological attachments of local leaders or the anticipated losses to those in comparison groups. Two authors of this paper (JH, PP) were involved in long discussions with an organization providing micro-finance to women in rural South Africa before agreement to use an RCT design was reached (example 1). In this case, researchers raised finances to support expansion of micro-finance programme to comparison clusters after the trial but within approximately the same time period had the trial not occurred. More generally, refinements to RCT designs, such as the stepped-wedge design, in which an intervention is progressively rolled out to all participating clusters and making use of other opportunities to include a randomization step, may be appropriate for trialling structural interventions. Some barriers may however be insurmountable, for example when there are administrative blocks to variable provision.

Trial statistical power is influenced by the numbers of both clusters and individuals within clusters. The feasibility of cluster-RCTs of structural interventions is therefore determined by the numbers of potential clusters available. Clusters should be the unit of social organization relevant to the intervention in question. Clusters included in a trial need to be sufficiently geographically or socially separated to prevent those in the comparison clusters being affected by the intervention (''contamination''). In some cases, only one cluster exists and therefore an RCT is impossible; for example, with global trade policies (example 1). Recruitment of clusters is likely to be easier where these are small units, such as villages, and more difficult where these are larger, such as when clusters need to be coherent entities in terms of welfare provision, legal jurisdiction, etc. (Des Jarlais, 2000; Sorensen et al., 1998).

Sometimes, a number of different cluster-based interventions may be being planned or implemented, all of which are hoped to have effects on the same health outcomes. In situations where only a limited number of clusters are suitable sites for pilot implementation, it is likely that the same clusters will feature in the intervention and comparison arms of the trials of two or more interventions. In such cases, the effects of each intervention cannot be distinguished. These problems are not particular to structural interventions but such interventions are particularly vulnerable to them because of their requirement that clusters are the appropriate unit of trial allocation. It should be noted that while these

issues present problems for the conduct of feasible RCTs they equally impede non-randomized designs.

Utility

In this section we consider the utility of using RCTs to evaluate structural interventions (as opposed to the utility of the interventions themselves). When do the benefits of undertaking a trial, in terms of new knowledge, outweigh the costs, in terms of money, time and harms experienced by participants? A trial may be considered ethical and feasible to undertake but this does not always mean it would be useful to undertake one, in terms of informing and influencing public health decision-making (McKinlay, 1993). Gains can arise both from positive and negative findings on the effectiveness and safety of interventions, since both results might beneficially influence policy, practice and resource allocation.

Proponents of structural interventions may decide that producing evidence of impact on HIV prevention from RCTs is essential in persuading those with the power to deploy such interventions, such as local politicians or foreign aid donors. This is particularly likely if such people are committed to preventing HIV infections but would not otherwise commit to supporting the interventions in question. This might occur for example where aid donors' focus is purely on HIV or health, so they would be uninterested in other intervention benefits such as poverty reduction (example 1). This might also occur where policy makers, in the absence of knowing about HIV-prevention effects, view interventions as ideologically unappealing (e.g. example 3). Such decision makers would require rigorous evidence of HIV-prevention effects in order to support the interventions in question. Evidence might be particularly important in settings where there is little existing recognition of the importance of socioeconomic and other 'upstream' influences on health (Leon & Walt, 2004).

However another possibility to consider, when deciding whether to trial structural HIV prevention interventions that are known to bring about other non-HIV benefits, is that trials might not actually be necessary to persuade decision makers to support the interventions. In such cases, RCTs might actually be a distraction from the task of deploying these interventions as rapidly and broadly as possible. Some decision makers might only require evidence of the 'plausibility' rather than the 'probability' of health effects in order to support the deployment of interventions for which there is already support (Habicht, Victora, & Vaughan, 1999). In such cases, non-RCT evidence may be regarded as sufficient, particularly when interventions are thought to bring other benefits and to have little possibility of being harmful. For example, UK HIV prevention organisations decided to develop anti-homophobia campaigns in the absence of evaluative evidence that such campaigns will influence gay men's HIV vulnerability, because the campaigns were regarded as beneficial for wider reasons and unlikely to bring about harm. In considering whether to implement structural interventions in such scenarios, proponents should remain alert to the possibility that interventions may not have the assumed effects, or that other interventions might achieve comparable effects more efficiently. A decision on whether to undertake an RCT of an intervention or not requires careful assessment on scientific and ethical grounds as discussed earlier.

A further matter is timing. RCTs may not be regarded as useful if they take too long to report outcomes, particularly with regard to infectious disease epidemics, such as with HIV, where decisions sometimes need to be made quickly. Some have suggested that structural interventions addressing upstream influences are likely to take longer than interventions addressing downstream influences to exert effects because more time is required for extended causal processes to operate (Berkman, 2004). This would seem likely in the case of the effects of household income in example 1, or of homophobia in example 2. However, the rapid deleterious impact of social upheaval on health indicators in eastern Europe suggests that the health effects of 'upstream' influences can occur more quickly than expected (Shkolnikov, McKee, & Leon, 2001). The question of whether RCTs can inform decisions with adequate timeliness will depend on: estimates of the timing of effects; whether valid proxy measures which manifest sooner are available; and epidemic characteristics determining the necessary speed of decision-making. Trials with insufficient length of follow up, like those with small sample sizes, might provide evidence subject to type-2 error perhaps leading to interventions being inappropriately abandoned.

RCTs should also be able to inform conclusions about the extent to which findings are likely to be reproduced in other settings. Structural interven-

tions addressing upstream influences, such as household poverty (example 1) are likely to be less generalisable than those addressing more downstream influences (such as police policy in example 3). This is because the causal chains linking interventions to outcomes involve more intervening factors that are likely to vary between settings. The potential for generalisability of interventions could be assessed in initial formative research to determine whether RCT evaluation is worth the investment. Such research would consider the scope for generalisability by exploring the process of intervention delivery and receipt and the contextual factors affecting this. Where formative research did suggest that an RCT might be worthwhile, trial outcome results could be used alongside process evaluation findings to develop recommendations concerning which types of settings the intervention might be deployable to, and what modifications might be required to maintain effectiveness within these (Grosskurth & Kumaranayake, 2003). Simulation modelling based on outcome and process results might be a useful tool in developing such recommendations.

When an RCT is not appropriate

When ethical and feasible, well-conducted RCTs of structural interventions for HIV prevention should provide the most accurate information possible on their effectiveness, other designs being weaker in their ability to minimise confounding and bias. In this paper, we have identified no factor that prevents structural interventions as a category being evaluated using this design. RCTs of structural interventions have successfully been undertaken in HIV prevention and other areas of sexual health. Including a randomisation step in planned deployments of such intervention might be more feasible than is currently accepted. However, we have identified situations in which it may not be ethical, feasible or useful to subject structural interventions that may have important HIV-prevention effects to evaluation using an RCT design. This is particularly the case: where structural interventions are known or widely thought to bring important benefits other than the intended HIV-preventative effects; where RCTs would require, but investigators fail to persuade, leaders of clusters to allow macro-social policies to be determined randomly; where the unit of social organisation addressed is so large that recruitment of adequate numbers of clusters is impossible; and where there is evidence that interventions exert effects across time periods over which RCTs cannot be maintained and where valid proxies do not exist. Similar situations have been identified for other types of interventions (Victora, Habicht, & Bryce, 2004).

We contend that the ability to evaluate an intervention with an RCT should be a less important factor in whether it is widely deployed than whether, according to the best evidence obtainable, it is a safe and effective intervention for HIV prevention. Therefore, in cases of interventions where use of the RCT design is not feasible, researchers and policy-makers should judge whether other designs can provide evidence of sufficient quality on safety and effectiveness and if so, use these to generate evidence to inform decisions about wider deployment. Evidence from non-RCT designs is most likely to be considered sufficient where existing observational evidence suggests there is little likelihood of the intervention bringing harm and where those making decisions about deployment are interested in the intervention's other benefits. In such cases simulation modelling to estimate the theoretical maximum impact of intervention strategies, alongside evidence of 'adequacy' of intervention delivery and 'plausibility' of HIV-preventative effects from field research may be sufficient to tip the balance of decisions (Habicht et al., 1999). Other evaluative designs that have been employed to establish causal effects of HIV prevention programmes include micro-econometric approaches, evaluations of dose-response models with respect to intervention access, and propensity-score matching (Coleman, 1999; Kelly, Kincaid, Parker, & Ntlabati, 2004; Magnani et al., 2004). A discussion of such alternative evaluation designs is beyond the scope of this paper. However, we additionally hope that recent advances in regulating the appropriate design, analysis and reporting of observational and non-randomised studies should inform judgments about the ability of non-RCT evaluations to inform questions of intervention design and deployment.

References

Angell, M. (1997). The ethics of clinical research in the third world. *Journal of Medical Ethics*, *337*, 847–849.

Berkman, L. F. (2004). Introduction: Seeing the forest and the trees–from observation to experiments in social epidemiology. *Epidemiological Review*, *26*, 2–6.

Blakenship, K. M., Bray, S. J., & Merson, M. H. (2000). Structural interventions in public health. *AIDS*, *14*(Suppl 1), S11–S21.

Bonell, C. P., Bennett, R., & Oakley, A. O. (2003). Sexual health should be subject to experimental evaluation. In J. Stephenson, J. Imrie, & C. Bonell (Eds.), *Effective sexual health interventions: Issues in experimental evaluation* (pp. 3–16). Oxford: Oxford University Press.

Coleman, B. E. (1999). The impact of group lending in Northeast Thailand. *Journal of Development Economics*, *60*, 105–141.

Des Jarlais, D. C. (2000). Structural interventions to reduce HIV transmission among injecting drug users. *AIDS*, *14*(S1), S41–S46.

Edwards, S. J., Braunholtz, D. A., Lilford, R. J., & Stevens, A. J. (1999). Ethical issues in the design and conduct of cluster randomized controlled trials. *British Medical Journal*, *318*(7195), 1407–1409.

Farmer, P., Connors, M., & Simmons, J. (1996). *Women, poverty and AIDS: Sex drugs and structural violence*. Monroe: Common Courage Press.

Friedman, L. M., Furberg, C. D., & DeMets, D. L. (1998). *Fundamentals of clinical trials* (3rd ed). New York: Springer.

Grosskurth, H., & Kumaranayake, L. (2003). Generalizability of trials and implementation of research into practice. In J. Stephenson, J. Imrie, & C. Bonell (Eds.), *Effective sexual health interventions: Issues in experimental evaluation* (pp. 169–190). Oxford: Oxford University Press.

Habicht, J. P., Victora, C. G., & Vaughan, J. P. (1999). Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact. *International Journal of Epidemiology*, *28*, 10–18.

Hayes, R. J., Alexander, N. D., Bennett, S., & Cousens, S. N. (2000). Design and analysis issues in cluster-randomized trials of interventions against infectious diseases. *Statistical Methods in Medical Research*, *9*(2), 95–116.

Johnson, S., & Rogaly, B. (1997). *Microfinance and poverty reduction*. London: Oxfam.

Kelly, K., Kincaid, L., Parker, W., & Ntlabati, P. (2004). Tackling the problem of non-equivalent exposed and unexposed groups in evaluation of mass media interventions: The evaluation of an entertainment-education television drama series. In XV *International AIDS Conference*. Bangkok, Thailand.

Leon, D., & Walt, G. (2004). Poverty, inequality and health in international perspective: A divided world? In D. Leon, & G. Walt (Eds.), *Poverty, inequality and health: An international perspective* (pp. 1–16). Oxford: Oxford University Press.

Magnani, R., Macintyre, K., Karim, A. M., Brown, L., Hutchinson, P., & Kaufman, C. (2004). The impact of life skills education on adolescent sexual risk behaviors in KwaZulu-Natal, South Africa. *Journal of Adolescent Health*, *36*(4), 289–304.

McKinlay, J. B. (1993). The promotion of health through planned sociopolitical change: Challenges for research and policy. *Social Science & Medicine*, *36*(2), 109–117.

McQueen, K.M., & Sugarman, J. (2003). HIV prevention trials network ethics guidance for research: HPTN Ethics Working Group.

Moher, D., Schultz, K. F., Altman, D. G. & CONSORT Group. (2001). The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials. *Lancet*, *357*, 1191–1194.

Oakley, A. (2000). *Experiments in knowing: Gender and method in the social sciences*. Cambridge: Polity.

Parker, R. G., Easton, D., & Klein, C. H. (2000). Structural barriers and facilitators in HIV prevention: A review of international research. *AIDS*, *14*(Suppl 1), S22–S32.

Patton, G., Bond, L., Butler, H., & Glover, S. (2003). Changing schools, changing health? Design and implementation of the gatehouse project. *Journal of Adolescent Health*, *33*(4), 231–239.

RADAR. (2002a). *Social interventions for HIV/AIDS. Intervention with microfinance for AIDS and gender equity. Evaluation monograph 1*. Acornhoek: Rural AIDS and Development Action Research Programme, School of Public Health, University of the Witwatersrand.

RADAR. (2002b). *Social interventions for HIV/AIDS. Intervention with microfinance for AIDS and gender equity. Evaluation monograph 2*. Acornhoek: Rural AIDS and Development Action Research Programme, School of Public Health, University of the Witwatersrand.

Rhodes, T., Singer, M., Bourgois, P., Friedman, S. R., & Strathdee, S. A. (2005). The social structural production of HIV risk among injecting drug users. *Social Science & Medicine*, *61*(5), 1026–1044.

Shkolnikov, V., McKee, M., & Leon, D. A. (2001). Changes in life expectancy in Russia in the mid-1990s. *Lancet*, *357*(9260), 917–921.

Sorensen, G., Emmons, K., Hunt, M. K., & Johnston, D. (1998). Implications of the results of community intervention trials. *Annual Review of Public Health*, *19*, 379–416.

Sumartojo, E. (2000). Structural factors in HIV prevention: concepts, examples, and implications for research. *AIDS*, *14*(Suppl 1), S3–S10.

Susser, M. (1996). Some principles in study design for preventing HIV transmission: Rigor or reality. *American Journal of Public Health*, *86*(12), 1713–1716.

Tawil, O., Verster, A., & O'Reilly, K,R. (1995). Enabling approaches for HIV/AIDS prevention: Can we modify the environment and minimize the risk? *AIDS*, *9*, 1299–1306.

UNAIDS. (2004). *Report on the global HIV/AIDS epidemic*. Geneva: UNAIDS.

Victora, C. G., Habicht, J. P., & Bryce, J. (2004). Evidence-based public health: Moving beyond randomized trials. *American Journal of Public Health*, *94*(3), 400–405.

Victora, C. G., Huttly, S. R., Fuchs, S. R., & Olinto, M. T. A. (1997). The role of conceptual frameworks in epidemiological analysis. *International Journal of Epidemiology*, *26*(1), 224–227.

Wight, D., & Obasi, A. (2003). Unpacking the 'black box': The importance of process data to explain outcome. In J. Stephenson, J. Imrie, & C. Bonell (Eds.), *Effective sexual health interventions: Issues in experimental evaluation* (pp. 151–166). Oxford: Oxford University Press.

Zierler, S., & Krieger, N. (1997). Reframing women's risk: Social inequalities and HIV infection. *Annual Review of Public Health*, *18*, 401–436.