



SAGA Working Paper
March 2005

**Scaling up HIV Voluntary Counseling and Testing in Africa:
What Can Evaluation Studies Tell Us About Potential Prevention
Impacts?**

Peter Glick
Cornell University

Strategies and Analysis for Growth and Access (SAGA) is a project of Cornell and Clark Atlanta Universities, funded by cooperative agreement #HFM-A-00-01-00132-00 with the United States Agency for International Development.



**Scaling up HIV Voluntary Counseling and Testing in Africa:
What Can Evaluation Studies Tell Us About Potential Prevention Impacts?**

Peter Glick
Cornell University
3M02 MVR Hall
Ithaca, NY 14853 USA
Phone 607-254-8782 (office)
607-257-4873 (home)
Fax 607-255-0178
pjpg4@cornell.edu

March 4, 2005

Author's Note: I would like to thank Steve Younger, David Sahn, and an anonymous referee for their comments.

Abstract

Although there is a widespread belief that scaling up HIV voluntary testing and counseling (VCT) programs in Africa will have large prevention benefits through reductions in risk behaviors, these claims are difficult to establish from existing evaluations of VCT. Considerations from behavioral models and the available data suggest that as VCT coverage expands marginal program effects are likely to decline due to changes in the degree of client selectivity, and that potential uptake among those at highest risk is uncertain. The paper also assesses two other common perceptions about VCT in Africa: that a policy of promoting couples-oriented VCT would be more successful than one emphasizing individual testing, and that VCT demand and prevention impacts will be enhanced where scaling up is accompanied by the provision of anti-retroviral drugs.

Keywords: HIV/AIDS, voluntary counseling and testing, selectivity models, program evaluation, behavior change

1. Introduction

In Africa, where AIDS has had its most devastating impacts and prevalence continues to rise in most countries, there is a crucial need for effective and feasible interventions for HIV risk behavior change. Increasingly, expansion of Voluntary HIV Testing and Counseling services (VCT) has been advocated as a central component of public health efforts to bring down HIV incidence through reductions in high-risk behaviors. Although only a small minority of adults in Africa are currently aware of their HIV status, many governments hope to change this by greatly expanding access to testing and counseling. VCT typically consists of a pre-test counseling session with a trained counselor, the serotest itself, and a post-test session in which individuals are counseled on behaviors to insure they remain uninfected (if they test negative) or avoid infecting others (if positive). Those testing positive are also provided emotional support, and directed to services to provide palliative care and other forms of support. Further, growing numbers of HIV positive people in Africa are expected to have access to antiretroviral therapies that can prolong their lives. Assessment of the need for ARVs naturally requires individuals to be tested.

Evaluations of VCT in Africa to date have focused on measuring the efficacy of VCT in changing risk behaviors of study participants or clients at VCT centers, usually in environments where the service was relatively new. They have not (as least directly) been concerned with what is the broader question of interest, the impacts on prevention behavior and ultimately HIV incidence in the target population for VCT when the intervention is expanded to become widely accessible to this population. That is, they have not evaluated what are variously called program “outcomes” or “effectiveness”.¹ Nevertheless, the results of these studies have been used by many researchers, international organizations, and AIDS policy advocates to argue that a massive scaling up of VCT services can be expected to contribute significantly to public efforts to change behavior and reduce HIV transmission (e.g., see WHO 2003a; Sweat et. al 2000). The purpose of this paper is to assess these claims, or more precisely, to assess the extent to which the available evidence on VCT in Africa can support them. By themselves, existing evaluations of VCT efficacy do not indicate either potential uptake of VCT or how behavioral responses to the program might change as coverage extends beyond initial study participants—two factors that are crucial determinants of the impacts of scaled up VCT programs on prevention behavior and the spread of HIV infection. This study therefore examines whether these gaps in knowledge can be filled by combining plausible conceptual frameworks for the determinants of the demand for HIV testing and its effects on behavior with data from evaluation studies themselves and other sources. It is argued that while some important clues about expanded program outcomes emerge, ultimately these outcomes remain very difficult to predict from the information that is available. Alternative evaluation approaches, notably community level evaluations, are required.

The rest of the paper is organized as follows. The next section provides a critical review of the evidence of VCT effects on behavior from evaluations conducted in African settings. Section 3 considers what can be determined about potential program outcomes from such studies. It sets out a simple framework indicating the information that existing VCT evaluations provide and the information they do not but which is also required to assess the impacts of scaling up. Alternative hypotheses are considered regarding the nature of self-selection of

individuals into VCT, which in turn will determine how behavioral responses may change as coverage expands. Data on testers are examined to evaluate one such hypothesis, that VCT attracts individuals with higher than average HIV risk. Information that may provide insights into potential VCT uptake—the other major requirement for predicting the impacts of a program expansion—is also considered. Section 4 shifts the focus from existing data and predictions to a discussion of appropriate evaluation strategies for understanding program outcomes. The next two sections address two important issues in the expansion of VCT. Section 5 considers whether an increasing emphasis on promoting couples oriented testing will be successful given apparently strong self-selection among couples in the use of VCT. Section 6 considers whether combining VCT expansion with the distribution of anti-retroviral therapies, as is likely to happen on a rapidly increasing scale in Africa, will enhance VCT demand and impacts. The final section summarizes.

2. Efficacy evaluations: does VCT change behavior?

A non-trivial number of published studies have investigated the effects of VCT in Africa, beginning with Kamenga et. al.'s (1991) study of Kinshasa, Zaire. Outcomes are almost exclusively self-reported behaviors such as the number of sexual partners and use of condoms. The most common study design is the one-group pretest and posttest design. Self-reported behaviors of VCT clients are recorded prior to and at some interval after the intervention and any change in behavior is attributed to the intervention. Few evaluations attempted to identify appropriate comparison groups, and only one study used a control group in the context of a randomized trial. In this study, conducted at sites in Tanzania and Kenya (as well as in Trinidad), individuals or couples who had been recruited into the study were randomized into VCT and basic 'health information' arms (VCT Efficacy Study Group 2000a).

Overall, this research provides evidence of some reduction in self-reported risk behaviors following HIV testing or VCT. Two main patterns emerge. The first is that risk-reducing behavior change tends to be larger among individuals who test positive than among those who test negative (Allen et. al. 1992b; VCT Efficacy Study Group 2000b; Van der Straten et. al. 1995; Lutalo et. al. 2000). This conforms to a general pattern observed for VCT elsewhere, including the U.S. (Weinhardt et. al. 1999; Wolitski et. al. 1997).² One contrasting finding comes from reviews of data from Uganda's AIC VCT program (UNAIDS 1999) which show that at 6-month follow-up, reported condom had risen strongly for both HIV positive and HIV negative clients. However, the share of HIV negative clients who were sexually active also increased.

The second common finding is that counseling of couples and/or partner testing appears to be effective at altering risk behavior, and more effective than individual testing and counseling when the two are compared (Allen et. al. 1992a,b; van der Straten et. al. 1995; Kamenga et. al. 1991; VCT Efficacy Study Group 2000b). In particular, serodiscordant partners (one HIV positive, the other negative) who test together adjust their behavior. 77% of Kamenga et. al.'s sample of serodiscordant couples reported using condoms during all episodes of sexual intercourse 18 months after testing and counseling compared with 5% before. Presumably these outcomes occur because couples counseling provides a means for broaching difficult subjects between partners. For women especially it may be crucial for risk reduction to have their male partner agree to testing and counseling, because in many African contexts men have control over

sexual decision-making (Van der Straten et. al. 1995; Ulin 1992). In some cases, individual testers also report risk reduction. Overall in the three sites of the multi-country VCT efficacy study, the percentage of individual testers reporting unprotected intercourse with non-primary partners with non-primary partners declined significantly from baseline, and significantly more than for the health information arm (35% vs. 13% reduction for men, 39% vs. 17% for women) (VCT Efficacy Study Group 2000b).

Not all African studies found reductions in risk behavior from testing or VCT. In particular, a number of analyses found little or no change in contraceptive use or pregnancy rates among women who tested HIV positive (Allen et. al. 1993; Ryder et. al. 1991; Heyward et. al. 1993; Temmerman et. al. 1990). Still, taken as a whole, the research to date gives the impression that VCT leads to reduced risk behaviors in certain groups who test—HIV positive individuals, and couples, particularly serodiscordant ones. They offer less reason for optimism about those who test negative. They suggest, therefore, that VCT will have significant impacts on the epidemic only if it is able to attract large numbers of HIV positive individuals, particularly those who are not yet ill, hence are still sexually active. To do this, the program—unless it is able to draw in such individuals highly disproportionately (and evidence discussed below suggests this is not assured)—must achieve broad coverage of the sexually active population to realize a significant public health impact.

Limitations of existing VCT evaluations

Most of these evaluations are subject to important limitations that may significantly reduce the usefulness of their findings. Almost unavoidably, given the sample sizes that would be required to use biomedical markers for transmission such as new HIV infections, almost all studies rely on self-reports of behavior, though the multi-site VCT study was able to use non-HIV sexually transmitted disease (STD) markers to confirm patterns in reported changes in behaviors. Follow-up periods in many cases are too short—often just several months—to gauge long-term impacts on behavior. Outcome measures are defined somewhat narrowly, which may limit the usefulness of the results to planners. In particular, none of the studies measure the effect of treatment on what health behavior change theories predict are key mediating variables, such as self-efficacy and behavior change knowledge or skills. This is despite the fact that in many cases (e.g., the multi-country study) the counseling strategies were said to have been explicitly informed by such theoretical frameworks and thus were designed to influence these mediating factors. It would be of interest to know whether interventions that explicitly incorporated constructs such as self-efficacy or behavior change skill acquisition are more effective, or whether the key benefit of VCT is simply that it provides information on one's status that can be used to adjust behavior. The more general point is that evaluations have not paid much attention to aspects of the counseling process, which clearly may differ across interventions and thus potentially determine efficacy.

However, the most serious limitations of this research arise from the study designs employed. First, as is well known, in single group pretest and posttest evaluations, changes in behavior may in part reflect general changes over time rather than the intervention. This is especially of concern with HIV-related interventions since knowledge about the epidemic was

probably growing during the duration of the studies. Equally important, given possible heterogeneity in the population with respect to (for example) motivation for behavior change, the results may reflect self-selection into VCT of individuals who are predisposed to make such changes. The likelihood of this is increased by the fact that in the contexts of almost all of these evaluations, VCT was still rare, hence somewhat novel. As discussed in the next section, this implies that the costs of using VCT (in this case referring to psychological or social costs rather than monetary costs) were probably quite high. Given this barrier, the VCT site or study may have drawn in individuals who were not representative of, and whose response to the intervention may differ from, the target population for the VCT program. As such, the implications for expanding the coverage of the program to a broader share of the target population are not clear.

For these reasons, a great deal of attention has been paid to the results of the one published randomized evaluation of VCT in Africa, the multi-center VCT efficacy study described above. The findings have been heralded in the public health literature and even in the popular press as proving the efficacy of VCT. The estimates have been used in simulations to demonstrate the cost effectiveness of an expanded VCT in saving lives (Sweat et. al 2000). However, even a rigorously conducted randomized study does not necessarily provide meaningful estimates of program efficacy for behavioral interventions such as VCT. Individuals are randomly assigned to intervention and control groups only after recruitment into the study. Recruitment itself is not random but relies necessarily on volunteers and these individuals may be unrepresentative of the target population in terms of motivation or other factors influencing responsiveness to the intervention. As just described, exactly this possibility—selectivity in participation that is correlated with responsiveness to the intervention—forms the basis of much of the criticism of uncontrolled studies. The randomized trial does insure stronger internal validity, meaning that it provides reliable estimates of the effect of VCT on those who volunteer, the ‘effect of treatment on the treated’. But external validity—the credibility of the findings as indicators of efficacy for the target population in general—may be weak.

This problem has long been recognized in the literature on evaluation (Kramer and Shapiro 1984; Heckman and Smith 1995). It suggests the need to interpret the results of the multi-country VCT study with a good deal more caution than has been the case. We would expect the threat to external validity to be large where acceptability of the intervention is low (self-selection into VCT is high) and where it is likely to be correlated with factors determining the response to the intervention (which behavioral models suggest may be the case, as discussed below). In the next section we try to assess the importance of these factors and what they might imply for program outcomes under an expansion.

3. Predicting Program Outcomes

We use some simple formal notation to indicate both the information that existing African VCT evaluations provide and the information they do not but which is also required to assess the impacts of a scaling up of VCT. Define Y as the behavior that is measured, say the reported number of sexual partners or the frequency of unprotected intercourse. The potential outcomes of individual i are Y_{1i} with the VCT program and Y_{0i} without it. The impact of the

program ('treatment effect') on i is therefore $\tau_i = Y_{1i} - Y_{0i}$. It is assumed to depend not just on whether i participates (is a VCT client) but also on individual characteristics X_i that can be observed (recorded in survey questionnaires) and on characteristics u_i that are likely to be unobserved. Thus we have $\tau_i = Y_{1i} - Y_{0i} | X_i, u_i, C_i = 1$ for participants ($C_i = 1$) and $\tau_i = Y_{1i} - Y_{0i} | X_i, u_i, C_i = 0$ for nonparticipants. First consider an evaluation using a randomly selected sample of VCT clients in an area where the service has been operating routinely and is known and available to anyone in the community or catchment area who wants it, i.e., there is no rationing. This seems to describe the context of some of the studies reviewed above. The VCT program outcome is defined as the average treatment effect for the target population, $E(\tau_i)$, calculated over those who participate and those who do not:

$$(1) \quad E(\tau_i) = E(Y_{1i} - Y_{0i} | X_i, u_i, C_i = 1) \cdot P_C(X_i, u_i, \tau) + E(Y_{1i} - Y_{0i} | X_i, u_i, C_i = 0) \cdot P_{NC}(X_i, u_i, \tau)$$

$E(\tau_i)$ thus is a weighted average of the effects on VCT clients and non-clients, with the weights equaling the shares of the target population using and not using the service, expressed equivalently as the probabilities P_C and P_{NC} of being or not being a client. These probabilities, like the treatment effects, are assumed to be functions of X_i and u_i as well as of the cost of using the service, denoted by τ . The second term is the average treatment effect conditional on not using the service multiplied by the probability of being in the non-client group. This term is zero unless there is an indirect effect of the intervention on those who do not actually participate.³ Thus as long as the program is already available to any who want it in the catchment area, the mean population effect is given by the first term alone. The evaluation study provides an estimate of the treatment effect of this group, $E(Y_{1i} - Y_{0i} | C_i = 1)$, though the standard one group pretest-posttest design means it is likely to be biased, as noted above. The study does not provide an estimate of P_C , the share of the target population that uses the VCT center. Information on the total number of clients visiting the center is presumably available, and it may be possible to come up with an estimate of the total adult population in the area served to provide a crude share figure. However, if the target population is defined to consist of adults specifically with high risk behavior, more detailed information on risk behaviors of the population at large from a population based survey is needed to define the target population and thus also the share of VCT clients in this population.

Planners are interested in an expansion of VCT services. If this involves only putting the same kind of VCT sites in other areas under similar conditions, the population outcomes in these areas, hence in the country as a whole, still just depends on the estimates of $E(Y_{1i} - Y_{0i} | C_i = 1)$ and P_C in eq. (1). However, the expansion of coverage under consideration may involve greater incentives or reduced costs, broadly defined. These can include not just lower financial costs (or for that matter, positive financial incentives such as transportation reimbursements for those who use the service) but also publicity campaigns to familiarize the population with VCT, to reduce fears or misconceptions about testing, and to address issues of stigma surrounding testing and a positive test result. When VCT is still rare, the psychological or social barriers to its use may be quite significant and the measures just described may reduce these costs, raising the acceptability of the program. At some lower cost τ_1 , additional people are induced to enter the program. The expected change in behavior in the target population is then (the X_i and u_i terms are suppressed to

avoid clutter but it should be understood that all treatment effects and probabilities are functions of these characteristics):

$$(2) \quad E(\Delta_i) = E(Y_{1i} - Y_{0i}|C_i=1) \cdot P_{C_i}(t) + E(Y_{1i} - Y_{0i}|C_i=1) \cdot P_{C_i}(t) \\ + E(Y_{1i} - Y_{0i}|C_i=0) \cdot P_{N_{C_i}}(t)$$

The first term on the right hand side is the same as before; those who used the service at the initial cost t_0 would also use it at t_1 .⁴ But the group that would not use VCT before is now composed of two subgroups, those who are induced to enter the program ($C_{1i}=1$) and those who remain non-participants ($C_{1i}=0$). To determine $E(\Delta_i)$, we need to know: $E(Y_{1i} - Y_{0i}|C_i=1)$ and P_C , as before; plus $E(Y_{1i} - Y_{0i}|C_i=1)$, the treatment effect among new participants (or the marginal treatment effect), and P_{C_1} , the share of the target population represented by this group. However, the evaluation study directly provides (potentially biased) information only on $E(Y_{1i} - Y_{0i}|C_i=1)$.

Next consider the case of a randomized trial conducted as a pilot study under somewhat different conditions. Here the service may not be available to all who want it; the researchers may simply stop recruiting once they have the desired number of individuals who meet the study criteria. Denote the mean outcome for the participants by $E(Y_{1i} - Y_{0i}|R_i=1)$, where $R_i=1$ indicates recruitment into the study. The target population mean change in behavior from the program if it were made readily available is given again by equation 1. However, given the recruitment setup just described, we know only the number of study volunteers, not the number of potential clients, so we are at even more of a disadvantage in trying to estimate P_C , the share of the target population that would use the service. Further, while randomization insures that $E(Y_{1i} - Y_{0i}|R_i=1)$ is a valid estimate of the effect of VCT on study participants, it may not be the same as $E(Y_{1i} - Y_{0i}|C_i=1)$, the mean program effect on VCT clients. However, in the case of the multi-country study discussed above it may be reasonable to assume that study participants were fairly representative of VCT clients as a whole, since the project seems to have recruited individuals specifically for VCT (VCT Efficacy Study Group 2000a). Then $E(Y_{1i} - Y_{0i}|R_i=1) \sim E(Y_{1i} - Y_{0i}|C_i=1)$ and we can write the population average treatment effect as

$$(3) \quad E(\Delta_i) = E(Y_{1i} - Y_{0i}|R_i=1) \cdot [P_{R_i}(t) + P_{N_{R_i}}(t)] + E(Y_{1i} - Y_{0i}|C_i=0) \cdot P_{N_{C_i}}(t)$$

where P_R is the share of the population recruited into the study. Eq (3) is the same as eq. 1 with P_C , the share of clients, divided into P_R and $P_{N_{R_i}}$, representing study participants and non-participants who would become clients if the service was available, respectively. Even after assuming that the mean treatment effect $E(Y_{1i} - Y_{0i}|R_i=1)$ on study participants is the same as it would be for VCT clients generally, we still lack information on the total number of clients $P_R + P_{N_{R_i}}$, since we know only the number of recruits P_R .

If as is likely the expansion of the program beyond the trials is associated with lowered costs the expected program outcome is similar to eq. (2):

$$(4) \quad E(Y_i) = E(Y_{1i} - Y_{0i} | R_i=1) \cdot P_{R_i}(1) + E(Y_{1i} - Y_{0i} | C_{1i}=1) \cdot P_{C_{1i}}(1) \\ + E(Y_{1i} - Y_{0i} | C_{1i}=0) \cdot P_{NC_{1i}}(1)$$

but here we lack information on all but $E(Y_{1i} - Y_{0i} | R_i=1)$, the effect of the treatment on study volunteers.

Is there any means of obtaining plausible estimates of the unknown quantities in this equation or equation (2) so as to be able to predict the program outcomes of scaled up VCT? One approach is to derive bounds for possible values for $E(Y_1)$ (see Manski 1995). To do this would in turn require hypothesizing bounds for the treatment effects of potential VCT clients under an expansion as well as for the number or share of new clients. Without more specific knowledge, especially on the share of the population that would use the service, this is likely to produce such a wide range of potential impacts as to be close to useless. However, it may be possible to do better by using additional data from the evaluation studies themselves or other sources, interpreted through plausible theoretical frameworks. These may help, for example, to sign the change in the treatment effect as the program expands. The additional information we are interested in would pertain to demand or potential demand for VCT services as well as how current VCT clients compare with non-clients in terms of characteristics that might condition responses to the program.

First we consider several alternative models of HIV testing behavior. They have in common the notion that the response to testing or counseling will differ among individuals based on characteristics (which may or may not be observed by researchers) that also affect the decision to participate in the program. One would begin by assuming that people are willing to try VCT if they expect the benefits to be greater than the costs.⁵ In addition to ending anxiety of not knowing one's HIV status, the value of the test result is that it provides information that allows one to adjust behavior; this information can be the test result itself or the knowledge obtained in counseling on safe sex practices or on negotiating safe sex with partners. With respect to costs, as noted, for several reasons the costs of testing, broadly defined, are likely to fall as VCT services expand. Therefore we would expect that as the service expands it will attract individuals who value it progressively less (Boozar and Philipson 2000).

Why would some people place a higher value on the information provided by VCT? First, there may be differences within the population with regard to motivation for behavior change. Those who seek HIV testing or counseling when VCT is new and 'cost' is high may be those who would be the most motivated to use the information they obtain to adjust their behavior; hence the test has the most value for them. A reduction in cost draws in people on the margin of participation who are less interested in adjusting behavior. This suggests that changes in behavior will be largest at first and then decline as the service expands, that is $E(Y_{1i} -$

$Y_{0i}|C_{1i}=1) < E(Y_{1i} - Y_{0i}|C_i=1)$.⁶ In contrast, Boozer and Philipson (2000) emphasize, not differential motivation, but differences in uncertainty about one's status. Two groups of people would have low uncertainty about their status: those who have engaged in so little risky behavior that they are almost certainly HIV negative and those who have engaged in so much that they feel they are very likely HIV positive. These individuals are able to adjust their behaviors more or less appropriately without having to test. On the other hand, for those who are more uncertain, pre-testing estimates of the probability of having been infected are less useful as guides to the appropriate behavior. These individuals have more to gain from testing and are more likely to make changes to behavior as a result.⁷ A VCT expansion that effectively lowers costs will bring in individuals who are less uncertain, hence are likely to change behavior by progressively smaller degrees. Hence again a decline in response at the margin would be expected.

A third possibility is that high risk individuals (rather than highly uncertain or highly motivated persons) value the service the most, because they are the ones who need to make the largest changes to behavior if they are to avoid becoming infected or infecting their partners. Thus they would have the most use for the counseling component of VCT that can assist them in developing strategies to reduce risk. Again, behavioral adjustments would decline (along with mean HIV risk) with program coverage under this scenario as lower costs draw in lower risk individuals who need to make relatively small changes to behavior. Importantly, however, where average levels of risk behavior and HIV prevalence rates are very high, many people would have room for significant behavior change so VCT coverage could extend a long way while still yielding large if diminishing treatment effects.

We next examine evidence from Africa on the risk and other characteristics of VCT participants. This may say something about the effect of VCT on those who test later in the growth of the program. Further, we still require an idea of the potential level of utilization of the program. Therefore we examine what is available on this also from the African evidence.

First, however, it is important to point out that treatment effects can change with an expansion of VCT coverage due to factors that do not involve individual heterogeneity and that are difficult to capture in surveys or individual-level evaluation studies. Significant among these factors are interactions among VCT participants or among participants and non-participants that influence how individuals respond to the program. They may, for example, make people more receptive to the behavior change messages they hear in counseling. People may be more willing to test or change behavior even without testing if they see that others around them are doing so. As VCT attains broader coverage such interactions or spillover effects will become more prevalent and if the effects are positive as just described, the effect of treatment will rise above that observed when coverage is limited. VCT advocates often note this possibility, referring to it in terms of increasing acceptability or changing norms of behavior. However, spillover effects may also be negative, reducing treatment effects as coverage expands. As individuals see larger numbers getting tested and presumably altering their behavior to reduce the risks of receiving and giving the virus, they may perceive their own risk of contracting HIV to be lower and feel less need to test at all or to alter their behavior after VCT.⁸ Finally, treatment effects can change because the treatment itself is heterogeneous. A particularly relevant example of this in a scaling

up scenario in African contexts is where shortages of resources such as skilled personnel begin to be felt, reducing the quality of the service hence also its efficacy in changing behavior.

Characteristics of VCT clients

It is often argued that VCT attracts individuals who are at high HIV risk relative to the general population. In the multi-country VCT study, baseline data confirm that clients were engaged in risky behaviors and had high levels of self-reported STD symptoms, though these characteristics were not compared to data for the population overall. 15% of men in the Kenya site and 12% of men in Tanzania tested HIV positive (Balmer et al. 2000; Sangiwa et. al. 2000). For women, rates of infection were very high (27% and 30%) and especially high for women not enrolling as part of a couple. In Uganda, in the relatively large-scale AIC program, 24% of individual testers were seropositive in 1997 though only 7% of couples tested positive (UNAIDS 1999; note that many of these couples came in for pre-marital testing, a group for which seropositivity is very low). Again the rates for women were much higher than for men: 26% vs. 14%. Among male factory workers in a study in Zimbabwe in 1993-97 (Machekano et. al. 2000), 20% of those who agreed to be tested (about one third of eligible workers) were seropositive.

How do these rates compare to the population overall in the study contexts? Seropositivity rates for men in the Kenya and Tanzania centers of the multi-site study are not higher than published estimates for the urban population from about the same time (World Bank 1999 Appendix Table 1). They are clearly much higher for women, however—a pattern also observed in a more recent Tanzania study (Chu et. al. 2004). This was also seen in Uganda, where female seropositivity rates were very high but male VCT clients had a seropositivity rate (14%) that was closer to the estimated population rate. A different study in Kenya (Forsythe et. al. 2002) found that HIV prevalence among VCT clients at two rural sites was generally lower than population estimates for the surrounding areas, but the opposite pattern obtained at an urban site. In the study of Zimbabwe factory workers, the men volunteering to test appear to have been at no higher risk—and quite possibly at lower risk—than the general population at that time in that location. Of course, when the population in general engages in high levels of risky behavior and HIV prevalence exceeds 15 or 20%, even reaching, and reducing risk behavior among, those at mean risk can significantly reduce new infections.

In a rare population-based study of VCT in Rakai, rural Uganda (Nyblade et. al. 2000) VCT services were offered to all individuals either in their homes or in a nearby clinic. There was no correlation of behavioral risk indicators and the decision to accept VCT services offered in either the home or a clinic, and for women HIV positive status (tested at baseline for the main Rakai STD/HIV study) was *negatively* associated with the decision to test. In urban Zambia, another population based survey offered VCT to all respondents (Fylkesnes and Siziya 2004). ‘Readiness’ for VCT (stating an interest in using the service offered) was associated with perceived HIV risk among 15-24 years olds and poor self-rated health (possibly perceived as an indicator of AIDS) among 24-49 year olds. However, there was no significant association of readiness and actual HIV status (as in Rakai, measured as part of the survey and separately from the VCT study). Therefore in these two cases in which there was direct information on HIV

status from probability samples, VCT did not attract those at relatively high risk as measured by serostatus itself.⁹

The evidence shows, therefore, that VCT attracts high-risk individuals (relative to the population) in some contexts but not in others, hence overall suggests the need for a nuanced view of the risk categories of testers.¹⁰ The much higher rates for female compared with male clients or volunteers at several sites—too large to be reflective of actual gender differences in the general adult population—is a clue that self-selection is very important, at least for women. The researchers for the Tanzania site suggested that the center may have attracted widows who believed their husbands had died of AIDS and wanted to see if they themselves had been infected (Sangiwa et. al. 2000). An alternative but not incompatible explanation is that the implicit costs of testing are higher for women. They probably have more to lose in terms of the stability of their partnerships from testing (if observed or discovered by their spouses) or from stigma generally. If they are less mobile, it may be harder for them to find ways to test discretely. If so, only women who place a very high value on the information provided may be willing to seek VCT, and as noted above, this may be individuals who are at particularly high risk.

Does the evidence just described provide insight into the marginal impacts on efficacy of a scaling up? Unfortunately, not very much. In some cases the evidence is consistent with idea of selection based on above-average HIV risk, but in other cases the situation is less clear. If it is not just high(est) risk individuals who use VCT, other factors, such as heterogeneity in motivation for behavior change or in uncertainty about one's status, may be at play. When selection occurs on these characteristics, changes in the effect of the program on behavior as VCT scales up are impossible to predict from existing data since unlike reported risk behaviors, we lack knowledge of the distribution of these characteristics in the population. Overall, the consideration of the evidence does not take us very far beyond *a priori* reasoning, which as noted generally points to a decline (at some unknown rate) in behavioral adjustments with an expansion.

Uptake of VCT

What about the potential level of utilization of VCT when availability increases—either $P_C(\cdot)$ in eq. (1) or $P_{C1}(\cdot)$ in eq. (2)? Some anecdotal evidence points to high demand at specific VCT centers, but others observers note that utilization of existing services in Africa has been disappointing (Fylkesnes 2000). To understand the demand for VCT we require data from population-based probability samples that include information on testing behavior. However, since VCT remains difficult to access for most people in Africa, information on testing from probability samples will usually not reveal the true demand for the service. Recent Demographic and Health Surveys in six African countries (Glick and Sahn 2004) indicate that a very large share – two-thirds or more – of individuals who do not know their status say they would like to get tested. However, the portion of adults who reported actually having been tested in these surveys was usually very low. For urban areas in several countries, notably Uganda and Zambia, rates of testing appear to be high (20% of adults 15-49) but elsewhere and in almost all rural samples rates fall below 15%. Given this, it is not possible to tell from the surveys how many of

those who say they would like to test in fact are actually reluctant to do so or are instead constrained by lack of access.

Therefore several recent studies that are population-based *and* were conducted in environments where VCT services were readily available are of particular interest. In the rural Rakai, Uganda study (Nyblade et. al. 2000), VCT services were offered to all individuals, who could choose to receive the service in their homes or at a nearby clinic. This policy experiment effectively reduced several kinds of costs associated with VCT to zero: the cost of the service itself, of transportation, and (by offering home testing and counseling) of risks to confidentiality associated with using a public facility.¹¹ Despite significant outreach, the level of demand in the initial year of the program (1995/5) was not very high—32% of women and 35% of men agreed to receive their test results. However, this jumped to 65% for both sexes in 1999/2000 (Matovu et. al. 2002), though it should be noted that this is conditional on having agreed in the first place to give a blood sample (78% of respondents).

In the urban Zambia study (Fylkesnes 2004), participants who had indicated ‘readiness’ to be tested were randomly assigned to receive VCT at either a clinic or in a setting chosen by the participant that included home counseling as one option. The mode of delivery mattered greatly: 56% of the group offered the choice of location used the service (most preferring home counseling) compared with only 12% in the clinic-only arm. However, even in the group with higher acceptability, only about 18% of the original sample went ahead with testing and counseling (i.e., both indicated readiness and actually used the service). This is lower than uptake in the rural Rakai study. However, as in Rakai, there is evidence of increasing acceptability of VCT. The acceptability of clinic-based VCT, while quite low (12%) was several times greater than found for a population-based study in the same setting conducted three years earlier.

In both studies, not only was VCT readily available to participants, costs were deliberately made very low in terms of time, money, and risk of loss of privacy. Therefore the results should provide an indication of the value of $P_{C1}(\cdot)$, the coverage than can be achieved by an expanded program with significantly reduced costs. In this sense the results are mixed with regard to the potential demand for, and hence the public health impact of, scaled up/low cost VCT. The rural Rakai results seem dramatic but the urban Zambia results less so. Both studies suggest that the mode of service delivery is a crucial determinant of acceptability, reflecting concerns about confidentiality and perhaps also a general lack of faith in local health service quality.¹² However, the home-based testing and counseling model evaluated in Rakai may not be feasible on a large scale given resource limitations. Also, this study environment was unusual because of the high local level of trust in the overall research project, which had been operating for years in the sample communities. Both studies also suggest, as does anecdotal evidence from various places around the continent, that acceptability of VCT in Africa has been rising over time.

To return to the question posed above, *a priori* considerations and the available evidence offer some clues, but limited ones, about program outcomes for VCT after efforts are made to scale up the service. Models of testing behavior provide reasons for expecting that the

responsiveness to the program will be lower among those who are brought in as coverage expands, but it is difficult to learn much more about this from the empirical evidence. However, some of the data, such as the very different rates of HIV infection of men and women at several sites, point to the complexity of the decision to seek testing and in so doing raise the likelihood of selectivity that could be strongly conditioning responsiveness to the program. As for uptake, the evidence remains rare but indicates that demand is responsive to changes that lower effective cost. The potential for achieving broad coverage may be significant when costs are very low in terms of monetary expense, access, and risks to privacy. In settings where this was the case, use of the service (or else readiness to use it) was not associated with higher HIV risk as measured by actual serostatus. What this and the relatively high coverage achieved might mean for the marginal and average treatment effects of the intervention is not known. Hence despite some insights, ultimately we are not able to assign reliable values for the unknown terms in the equations for the outcomes of expanded VCT programs.

4. Future Directions for VCT Evaluations

As noted, randomized individual level efficacy analyses on study participants can have high internal validity but low external validity, and they do not provide information on uptake.¹³ Hence further VCT evaluations using this approach are unlikely to fill the gaps in knowledge identified above. More promising is the use of appropriate statistical methods on non-experimental data. Plausible behavioral models discussed in the last section suggest the existence of individual heterogeneity in response to the intervention that is also related to the decision to participate in VCT. Although heterogeneity may be captured by observable factors, i.e., characteristics recorded in survey questionnaires, more generally it will also reflect factors that are not measured by surveys.¹⁴ Because of this, instrumental variable methods are required. Research in recent years has focused on the appropriate interpretation of IV estimates when heterogeneity in the effects of programs exists and is related to participation. This literature has shown (Imbens and Angrist 1994; Heckman 2001) that under these conditions standard IV estimates only measure the effect of treatment on those who are induced to enter the program by the change in the instrumental variable employed; hence they do not yield either the average treatment effect or marginal treatment effects of a particular policy unless the policy exactly corresponds to the change in the instrumental variable (e.g., price) observed in the existing data. Recently, however, Heckman and colleagues (Heckman and Vytlačil forthcoming; Carneio, Heckman and Vytlačil 2003) have proposed more general estimators that are not subject to this limitation, permitting estimation of the effects of relevant potential policies that change the level of program participation.

These tools hold promise for addressing the questions motivating this paper. Still, a stumbling block is the possibility of social interactions among individuals and general equilibrium effects noted earlier, which are problematic for causal inferences in analysis conducted at the individual level.¹⁵ A strategy that is better placed to deal with these concerns is community level randomization. Policy experiments using community randomization are increasingly being applied to health and education interventions in developing countries, including interventions to reduce STDS and HIV (Miguel and Kremer 2003; Wawer et. al. 1998). Program outcomes are measured as the difference in mean outcomes in intervention and

control communities, estimated from probability samples that included treated and non-treated individuals. Randomization over communities rather than individuals would provide direct estimates of average program outcomes in the target population for VCT, incorporating both uptake and efficacy and sidestepping the problem of self-selection into treatment. These outcome measures capture the effects of interactions among agents that may influence outcomes among program participants as well as among those who do not participate, and also will capture community level general equilibrium effects. There are of course drawbacks as well, notably with respect to organizational complexity and cost, especially given the number of control and treatment communities that is typically required. The approach is difficult to carry out within urban areas since individuals may more easily travel from control to treatment communities to seek the intervention, contaminating the experiment. It is also probably too blunt and costly to be used to fine tune counseling strategies.

The design of community level experiments should recognize the likelihood that HIV testing will eventually be available in one form or another on a wide scale. Therefore it would probably not be very useful, and may be unethical, to have control communities receiving no testing of any kind. Rather, study arms should receive different modes of service delivery. This can answer important questions about both public health impacts and public sector costs of different kinds of VCT options that are likely to differ greatly in terms of the resources required—a highly germane issue for Africa in view of severe financial and personnel resource shortfalls.

For example, high coverage was achieved in Rakai through in-home counseling services, but this is expensive. A more feasible option in many settings would be to set up community-based services using trained non-professionals, or to provide mobile testing sites. Communities featuring these delivery modes could be compared to communities receiving the service in a standard health clinic setting. Another, quite controversial, policy option currently being debated is to expand the numbers getting testing by conducting routine or ‘mandatory’ testing of anyone entering a health center or hospital (see UNAIDS 2004 and Holbrooke and Furman 2004 for opposing views). An experiment should be feasible to set up if the health authority was willing to institute mandatory testing in several health districts in a pilot study.

Finally, in the coming years access to life-prolonging anti-retroviral (ARV) drugs will become a reality for many Africans with HIV/AIDS. As discussed further below, this may have important implications for the demand for testing and its effects on behavior. Given resource and logistical constraints, the introduction of testing sites providing the drugs undoubtedly will be staggered (in fact this is exactly what is happening now): the drugs will initially be available in certain areas and not others. The latter can form natural ‘late treatment’ controls during the period for which ARVs are still unavailable in them. Since these communities would be receiving the extant standard of care, the ethical concerns that might arise in the experiment are avoided or reduced.¹⁶

5. Implications for expansion of couples vs. individual testing

Proponents of scaling up VCT argue not just for an expansion of the service but an increasing orientation toward testing of couples. This is based on the evidence cited above that

behavior change is greater when people test in couples than when testing as individuals, particularly when the partners are serodiscordant. Therefore VCT programs should make special efforts to attract couples (UNAIDS 2001; Painter 2002). However, the impacts on HIV transmission of such a strategy will depend on how successful it is at bringing couples in to test, as well as on whether and how the effects of VCT on risk behaviors of couples changes with the expansion of coverage.

It is difficult to assess the demand for couples or ‘joint partner’ testing simply by looking at the composition of clients or study participants in existing VCT evaluations, because many of the interventions studied went out of their way to recruit couples or (as in the VCT efficacy study) insured by design a given mix of couples and individuals. However, a number of studies point to a strong reluctance on the part of many to test with their partners. In the Zimbabwe study of factory workers (Machekano et. al. 2000), only 7% of the men who tested brought in their partners for testing, despite efforts to make this convenient for them. A study of VCT sites in several Southern African countries reports that typically few couples came in for testing, despite efforts to encourage couples to test (UNAIDS 2001). In Lusaka, couples testing found very few takers (Baggaley et. al. 1997). These impressions of low demand sit uneasily with the evidently high responsiveness to the intervention of couples who do come in. Is the latter a reflection of particularly strong self-selection?

One barrier to more couples testing may be that many African women who would like to have themselves and their partners tested lack the power to get their partners to go along. Suggesting a joint test may indicate that they harbor suspicions that their partners have outside relationships, which itself could have negative repercussions. For a partner (male or female) who has outside relationships, (Glick 2004) notes that the costs of joint testing may be higher than testing singly due to information asymmetries between partners. Agreeing to or suggesting joint testing may be tantamount to revealing an external relationship not previously known to the other partner. With joint (or more generally, ‘partner-observed’) testing, this cost is always incurred by the first partner, even if the test turns out to be negative. Hence even an individual who is altruistic in the sense of being willing to inform his partner if the result was positive (and thus also to acknowledge the infidelity) would all things equal prefer to test without the partner’s knowledge: if he tests positive, the costs to him are the same as with joint testing (reveal HIV status and the infidelity), but if the result is negative, the cost of revealing the outside relationship need not be incurred, unlike with a negative result under joint testing.

Whatever the reasons, many people obviously perceive the costs to testing with their partners to be very high. Only individuals and couples who are particularly committed to cooperate in behavior modification are likely to accept joint testing. To most observers the solution is to reduce the costs though efforts to encourage better communication and cooperation regarding behavior change between partners (UNAIDS 2001). Still, the available evidence of weak demand for couples testing points to strong self-selection. Planners should be cautious in concluding that expanding VCT programs to include a greater share of couples, for example by offering greater subsidies to couples to test, will lead to responses similar to those recorded in existing studies.

6. Implications of scaling up while making antiretroviral therapy accessible

Though the potential for meeting the total need is far from certain, it is clear that in coming years large numbers of Africans with HIV/AIDS will be provided with life-prolonging anti-retroviral drug therapies. A frequently cited secondary benefit of providing the drugs is that it will increase demand for HIV testing—after all, testing is the gateway to the therapy—thus enhancing the preventative impacts of testing or VCT (Maotti 2002). An expansion of VCT programs that occurs in the context of a scaling up of ARV distribution may therefore yield demand and prevention outcomes that are quite different than in the absence of ARVs. Here I consider what the available evidence and *a priori* considerations can tell us about these potential differences.

Given that in Africa the distribution of free or subsidized ARVs has only recently begun on a limited scale, the VCT evaluations reviewed above are not going to be very informative on this issue. Conceptually, however, it seems that outcomes with respect to the demand for testing would depend significantly on the criteria used for determining who is eligible for treatment. In most existing pilot sites in Africa (in particular, in South Africa, Botswana, and Uganda), as in developed countries and in accordance with recent WHO guidelines (WHO 2003b), decisions to start ARV therapy are based on an individual's T-cell (CD4) count; even asymptomatic HIV positive persons with sufficiently low CD4 counts (below 200 cells/mm³) are advised to start ARV treatment. When it comes to scaling up ARV therapy this may well prove impractical because of the costs and level of sophistication of the CD4 tests. Until cheaper CD4 testing becomes available, the only feasible basis for deciding to start therapy in HIV-positive individuals on a wide scale may be whether the person has developed symptomatic AIDS. Perhaps even more importantly, the demand for free or subsidized AIDS drugs will almost certainly outstrip supply for a long time to come. This means that they must be rationed, and the most likely way this will be done is by restricting distribution to symptomatic AIDS patients, among whom the disease has progressed the most and for whom the benefits of the therapy are the greatest.

If either of these factors prevail, only those who have already become ill with AIDS would be eligible for treatment. Using the basic conceptual framework of Section 3, ARV availability under this eligibility criterion would be predicted to have heterogeneous effects on the incentive to test, increasing it for some but decreasing it for others. The value of testing would clearly rise for those who are experiencing illnesses that they or their health care practitioners think may be AIDS-related. Such individuals will receive the drugs if they test positive. The subsequent improvement in their health will tend to make them more sexually active but also less infectious than before, so the implications for the spread of the virus is not clear. But for those who are not (or not yet) ill there is no additional incentive to test. Access to drugs that can be obtained only after one falls ill does not raise the value of the information that the test provides for those who are not ill.¹⁷ However, from a prevention point of view it is precisely this group—or, given the evidence on behavioral responses by serostatus, those among them who are HIV positive—that should be tested. In fact, the value of testing would fall for this group, all things equal. By effectively reducing the difference in life expectancy with and without HIV/AIDS, ARV availability reduces the cost associated with not knowing one's status: one may be taking a risk by having unprotected sex, but the cost, in terms of possible reduction

in one's life expectancy or that of a partner one might infect, is lower. For the same reason, risk behavior adjustments of those who do test may be smaller.

From a prevention standpoint, if making ARV therapy available does not increase the demand for testing among sexually active individuals who are either HIV negative or (especially) pre-symptomatic HIV positive, it is unlikely to make VCT any more effective at prevention than it would be otherwise. This suggests a potential tradeoff between the objectives of allocating scarce drugs only to those who need it most, on the one hand, and of increasing the demand for testing by offering drugs to at least some individuals who are infected but not yet sick on the other.

The only evidence so far on any of this consists essentially of informal reports from pilot ARV sites (AIDSMAP 2004; Mpiima S et al. 2003, WHO 2003c). Apparently confounding the *a priori* considerations just outlined, they suggest a surge in the numbers of individuals seeking testing that is clearly related to the introduction of the drugs. The precise reasons for this are not clear but are important to learn. Clearly the most favorable explanation from a public health perspective is that the possibility of treatment, even down the line, reduces the stigma and fear of learning one's status and encourages even apparently healthy individuals to come forward.¹⁸ Another is that in most of these pilot sites, even asymptomatic people have an incentive to test since they can still receive the drugs if their CD4 count meets the criteria. A third possibility is that individuals are not yet aware of the criteria for treatment and believe they can receive the treatment even if they are not yet ill. Finally, a fourth possibility is that most people coming to test are ill with AIDS-related symptoms, so that demand has not really increased among the apparently healthy. There is a suggestion that this characterizes the Botswana experience (UNAIDS 2003 p. 13). To be able to predict whether the apparent strongly positive impact of ARV availability on VCT uptake will be replicated as the programs are scaled up, it is important to know the composition by HIV and morbidity status of those demanding testing at these sites, their expectations concerning access to the drugs, and the criteria that will be used to determine eligibility when the program is expanded and supply constraints begin to be felt. It should also be clear that there is a need for new rigorous assessments of VCT demand and impacts in settings where ARVs are available.

7. Conclusions

This paper has emphasized the difficulty of inferring program outcomes from existing evaluations of VCT efficacy in Africa when potential uptake or acceptability is not known and behavioral responses at the margin may change significantly as the availability of VCT increases. Because of the importance of these factors, reductions in risk behaviors reported in these studies may have limited relevance for inferring the outcomes from a scaled up program that attempts to achieve broad coverage. Even individual level randomized trials may have low external validity, hence may not provide meaningful estimates of behavioral outcomes under an expanded program. A key problem is that expansion is likely to be associated with significantly reduced costs (broadly defined) to using the service, and this will draw in new participants who are likely to respond differently than existing ones who presumably value the service more highly. Simple but plausible models of testing behavior raise the possibility that treatment effects will become smaller as coverage expands. The strongest changes in behavior in VCT evaluations have been

observed for (serodiscordant) partners who test together. However, this is also the case where self-selection in program participation seems to operate most strongly, suggesting the need for caution in predicting the prevention impacts of a strategy of targeting VCT to couples.

Ultimately a review of existing evaluations and consideration of other available data (such as on uptake) leaves us unable to reliably predict program outcomes. If this conclusion is unsurprising, it serves as a reminder of the need to be humble about what we can say about VCT prevention impacts based on the information we have. Future evaluation efforts would benefit from either more sophisticated non-experimental methods or, optimally, community randomized designs to avoid the limitations inherent in clinical or simple observational studies. VCT (or at least, testing) on a significantly expanded scale is inevitable in most African countries, and can be justified by a number of reasons other than simply prevention. For policy makers in these countries, the question is now how to achieve the greatest impact with limited resources. Evaluation efforts could therefore be profitably directed at comparing outcomes from different modes of service delivery, which may differ significantly in terms of costs to governments or donors.

Future efforts should also be directed at assessing the demand for and impacts of VCT in contexts where anti-retroviral drug therapies are made available. The implications of drug availability for the demand for testing among individuals who are not ill but are at risk (of giving or getting the virus) may depend significantly on the criteria adopted for determining eligibility for the drugs. Therefore careful evaluation of the effects on VCT demand—and its impacts—of expanding access to drug therapies under specific eligibility rules is needed.

References

- AIDSMAP. 2004. Integrating HIV prevention and treatment. HIV & AIDS Treatment in Practice #30. <http://www.aidsmap.com>
- Allen S., J. Tice, P. Van de Perre, A. Serufilira, E. Hudes, and F. Nsengumuremyi, et al. 1992a. Effect of serotesting with counseling on condom use and seroconversion among HIV discordant couples in Africa. *British Medical Journal* 304:1605-9.
- Allen S., A. Serufilira, J. Bogaerts, P. Van de Perre, F. Nsengumuremyi, and C. Lindan, et al. 1992b. Confidential HIV testing and condom promotion in Africa: impact on HIV and gonorrhea rates. *Journal of the American Medical Association* 268 23:3338-3343.
- Allen S., A. Serufilira, V. Gruber, S. Kegeles, P. Van de Perre, M. Carael, and T.J. Coates. 1993. Pregnancy and contraceptive use among urban Rwandan women after HIV counseling and testing. *American Journal of Public Health* 83:705-710.
- Baggaley R., F. Drobniewski, A. Pozniak, D. Chipanta, M. Tembo, and P. Godfrey-Faussett. 1997. Knowledge and attitudes to HIV and AIDS and sexual practices among university students in Lusaka, Zambia and London, England: are they so different? *Journal of the Royal Society of Health* 117:88-94.
- Balmer, D.H., O.A. Grinstead, F. Kihuho, S.E. Gregorich, M.D. Sweat, and M.C. Kamenga, et al. 2000. Characteristics of individuals and couples seeking HIV-1 prevention services in Nairobi, Kenya: the voluntary HIV-1 counseling and testing efficacy study. *AIDS and Behavior* 4:115-23.
- Boozer, M. A. and T. J. Philipson. 2000. The Impact of Public Testing for Human Immunodeficiency Virus. *Journal of Human Resources* 35(3:419-446.
- CADRE. 2002. On the move: the response of public transport commuters to HIV/AIDS in South Africa. Centre for AIDS Development, Research and Evaluation, and Department of Health.
- Carneiro, P., J. Heckman, and E. Vytlačil. 2003. Understanding What Instrumental Variables Estimate-Estimating Marginal and Average Returns to Education. Working Paper, Stanford University.
- Cartoux M., P. Msellati, and N. Meda, et al. 1998. Attitude of pregnant women towards HIV testing in Abidjan, Cote d'Ivoire and Bobo-Dioulasso, Burkina Faso. *AIDS* 12: 2337-2344

- Chu, H. Y., R.B. Oenga, D.K. Itemba, A. Mgonga, S. Mtwewe, J.A. Crump, J.A. Bartlett, J.F. Shao, and N.M. Thielman. 2004. Testing Strategies and Client Characteristics at a Voluntary Counseling and Testing Center in Moshi, Tanzania. The 11th Conference on Retroviruses and Opportunistic Infections, San Francisco, California.
- Forsythe, S., G. Arthur, G. Ngatia, R. Mutemi, J. Odhiambo, and C. Gilks. 2002. Assessing the cost and willingness to pay for voluntary HIV counseling and testing in Kenya. *Health Policy and Planning* 17 (2):187-195.
- Fylkesnes, Knut. 2000. Consent for HIV Counseling and Testing. *Lancet* 356 (Suppl December):s. s43.
- Fylkesnes, K. and S.A. Siziya. 2004. A randomised trial on acceptability of voluntary HIV counselling and testing. *Tropical Medicine & International* vol. 9, iss. 5, pp. 566-572(7).
- Glick, Peter. 2004. Interpreting evidence on HIV testing behavioral outcomes in Africa. Draft. Cornell University.
- Glick, Peter, and D. Sahn. 2004. Changes in HIV/AIDS Knowledge And Testing Behavior In Africa: How Much and for Whom? Cornell University Food and Nutrition Policy Program Working Paper 173, Ithaca , NY.
<http://www.cfnpp.cornell.edu/info/wp173.html>
- Heckman, James J. and J.A. Smith. 1995. Assessing the Case for Social Experiments. *Journal of Economic Perspectives* 9 (2) 85-110.
- Heckman, James. 2001. Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture. *Journal of Political Economy* vol. 109(4), pp.673-748.
- Heckman, J. and E. Vytlacil. Forthcoming. Structural Equations, Treatment, Effects and Econometric Policy Evaluation. *Econometrica*.
- Heyward W., V. Makizayi, V.L. Batter, M. Malulu, N. Mbuyi, L. Mbu, M.F. St Louis, M. Kamenga, and R.W. Ryder. 1993. Impact of HIV counseling and testing among child-bearing women in Kinshasa, Zaire. *AIDS* vol. 7: 1633-1637.
- Holbrooke, Richard and R. Furman 2004. "A Global Battle's Missing Weapon." *New York Times*, February 10, 2004.
- Kamega M., R. Ryder and M. Jingu, et al. 1991. Evidence of marked sexual behavior change associated with low HIV-1 seroconversion in 149 married couples with discordant HIV-1 serostatus: experience at an HIV counseling centre in Zaire. *AIDS* 5:61-67.
- Kramer, M.M. and S.H. Shapiro. 1984. Scientific challenges in the application of randomised trials. *Journal of the American Medical Association* 252, 2739-45.

- Lutalo, T., M. Kidugavu, and M. Wawer. 2000. Contraceptive use and HIV counseling and testing in rural Rakai district, SW Uganda. Presented at the 13th International Conference on HIV/AIDS Durban, South Africa (Abstract C246).
- Machekano R., W. McFarland, E. Hudes, M.T. Bassett, M.T. Mbizvo, D. Katzenstein. 2000. Correlates of HIV Test Results Seeking and Utilization of Partner Counseling Services in a Cohort of Male Factory Workers in Zimbabwe. *AIDS and Behavior* 41: 63-70.
- Manski, Charles. 1995. Learning about Social Programs from Experiments with Random Assignment of Treatments. Discussion Paper no. 1061-95 Institute for Research on Poverty, University of Wisconsin-Madison.
- Miguel, E. and M. Kremer. 2003. Health Behavior and the Design of Public Health Programs: Evidence from Randomized Evaluations. Department of Economics, University of California at Berkeley. <http://post.economics.harvard.edu/faculty/kremer/papers.html>
- Matovu J.K.B., G. Kigozi, F. Nalugoda, F. Wabwire-Mangen, and R.H. Gray. 2002. The Rakai Project counseling program experience. *Tropical Medicine & International Health* vol. 7, iss. 12, pp. 1064-1067(4).
- Mpiima S, et al. 2003. Increased demand for VCT services driven by introduction of HAART in Masaka District, Uganda. Poster presentation to the Second IAS Conference on HIV Pathogenesis and Treatment, Paris, July 13-17, 2003.
- Nyblade, L., R. Gray, F. Makumbi, T. Lutalo, J. Menken, M. Wawer, N. Sewankambo, and D. Serwadda. 2000. HIV risk characteristics and participation in voluntary counseling and testing in rural Rakai district, Uganda. Presented at the 13th International Conference on HIV/AIDS Durban, South Africa (WeOrD63).
- Painter, T. 2001. Voluntary counseling and testing for couples: a high-leverage intervention for HIV/AIDS prevention in sub-Saharan Africa. *Social Science and Medicine* 53:1397-1411.
- Rubin, D. B. 1990. Formal Models of Statistical Inference for Causal Effects. *Journal of Statistical Planning and Inference* vol. 25: 279-292.
- Ryder R., V. Batter, and M. Nsuami. 1991. Fertility rates in 238 HIV-1 seropositive women in Zaire followed for 3 years post-partum. *AIDS* 5:1 521-27.
- Sangiwa M.G., O.A. Grinstead, M. Hogan, D. Mwakagile, J.Z.J. Killewo, S.E. Gregorich, M.C. Kamenga, and M.D. Sweat. 2000. Characteristics of individuals and couples seeking HIV-1 prevention services in Dar Es Salaam, Tanzania the voluntary HIV-1 counseling and testing efficacy study. *AIDS and Behavior* 4:1 25-33.
- Sweat M., S. Gregorich, G. Snagiwa, C. Furlonge, D. Balmer, C. Kamenga, O. Grinstead, and T. Coates. 2000. Cost-effectiveness of voluntary HIV-1 counseling and testing in

- reducing sexual transmission of HIV-1 in the United Republic of Tanzania and Kenya. *Lancet* 356:113-21.
- Temmerman M., S. Moses, D. Kiragu, S. Fusallah, I.A. Wamola, and P. Piot. 1990. Impact of single session post-partum counseling of HIV infected women on their subsequent reproductive behavior. *AIDS Care* 247-252.
- Ulin, P. 1992. African women and AIDS: Negotiating behavioral change *Social Science & Medicine* vol. 34 (1): 63-73.
- UNAIDS. 2003. Stepping back from the edge: The pursuit of antiretroviral therapy in Botswana, South Africa and Uganda. UNAIDS Best Practice Collection.
- UNAIDS. 2001. HIV Voluntary Counseling and Testing: A Gateway to Prevention and Care. UNAIDS Best Practice Collection.
- UNAIDS. 1999. Knowledge is power: voluntary HIV counseling and testing in Uganda. UNAIDS Best Practice Collection.
- UNAIDS. 2004. Policy Statement on HIV testing. Information Note.
- van der Straten A., R. King, O. Grinstead, A. Serufilira, and S. Allen. 1995. Couple communication, sexual coercion and HIV risk reduction in Kigali, Rwanda. *AIDS* vol. 9:8 935-44.
- VCT Efficacy Study Group. 2000a. The Voluntary HIV-1 Counseling and Testing Efficacy Study: Design and Methods. *AIDS and Behavior* vol. 4(1), 5-14.
- VCT Efficacy Study Group. 2000b. Efficacy of voluntary HIV-1 counseling and testing in individuals and couples in Kenya, the United Republic of Tanzania and Trinidad: a randomized trial. *Lancet* 356:103-12.
- Wawer M.J., N.K. Sewankambo, D. Serwadda, T.C. Quinn, and L.A. Paxton. 1998. A randomized, community trial of intensive STD control for AIDS prevention, Rakai District, Uganda. *AIDS* (12): 1211-1225.
- Weinhardt, L.S., M.P. Carey, B.T. Johnson, and N.L. Bickham. 1999. Effects of HIV counseling and testing on sexual risk behavior: A meta-analytic review of published research, 1985-1997. *American Journal of Public Health* vol. 89(9): 1397-1405.
- WHO. 2003a. The Right to Know: New Approaches to HIV Testing and Counseling. <http://www.who.int/hiv/pub/vct/pub34/en/>
- WHO 2003b. Scaling up anti-retroviral therapy in resource limited settings: treatment guidelines for a public health approach. http://www.who.int/hiv/pub/prev_care/en/WHO_ARV_Guidelines_Update.pdf

- WHO. 2003c. Antiretroviral Therapy in Primary Health Care: Experience of the Khayelitsha Programme in South Africa Case Study http://www.who.int/hiv/pub/prev_care/pub38/en/
- Wolitski, R.J., R.J. MacGowan, D.L. Higgins, and C.M. Jorgensen. 1997. The effects of HIV counseling and testing on risk-related practices and help-seeking behavior. *AIDS Education and Prevention* 9(Suppl. B) 52-67.
- World Bank. 1999. *Confronting AIDS: Public Priorities in a Global Epidemic* Oxford University Press.

Endnotes

¹Terminology is not always consistent in the evaluation literature, so it should be emphasized that “program outcomes” as used here refers to the target population impacts of programs, not to the effects of VCT observed among study participants, which is what the existing VCT efficacy studies measure. In most discussions in African contexts the target population for expanded VCT includes all sexually active adults.

²Simple economic behavioral models suggest that *negative* testers should take greater steps to protect themselves, since HIV negative persons have a longer life expectancy, hence more to lose from risky behaviors that could lead to infection. Glick (2004) suggests two reasons for a lack of reductions in their risk behaviors: first, a negative test result leads individuals to revise downward the estimated probability that their partners(s) are HIV positive, hence also their risk of becoming infected by their partners; second, ‘altruistic’ individuals who test negative may increase or at least not reduce risky behavior because they realize they are less of a danger to others.

³ We make this assumption here but as discussed below such indirect effects may exist.

⁴ They may also use the service more intensively, that is, get tested more than once or more frequently. This complication is ignored here.

⁵ This may smack of excessively economic logic, but as already seen, ‘costs’ can be considered broadly enough to encompass non-economic factors such as stigma or fear and the same is true of benefits.

⁶ On the other hand, it is possible that more highly motivated or knowledgeable people have already made significant adjustments to their behavior prior to coming in for VCT so have fewer changes to make.

⁷ Among those who ex ante condition their behavior on an (accurately estimated) infection probability that is close to one or the other extreme, say 0.10 or .85, behavior on average is already close to what it would be if actual status was known. The greatest uncertainty is for those whose expected probability falls midway between the extremes of 0 and 1: for example, for those whose estimated probability of being HIV positive is 0.5, the mean difference between estimated and actual status (which must be 0 or 1) will be 0.5. These individuals have the most to gain from testing and will make on average the largest adjustments to behavior following testing.

⁸ This can be interpreted as a general equilibrium effect: by reducing the average risks of infection and loss of future utility due to early death, increasing numbers of people testing and practicing safe sex lowers the ‘price’ to an individual of having more partners or unprotected sex.

⁹ Among adolescents and young adults especially, high-risk behaviors may not yet have resulted in high rates of HIV infection, so HIV status itself may actually be a less informative measure of HIV risk than behaviors. On the other hand, if as some the evidence indicates, only HIV positive individuals change their behavior after VCT, then HIV status is the more relevant indicator of the target group for the intervention for any age category.

¹⁰ The fact that until fairly recently same day test results were not available provides additional evidence on the tenuousness of the link between HIV risk and the demand for testing. A number of studies from Africa (Cartoux et. al. 1998; Temmerman et. al. 1995) showed that individuals who tested HIV positive were less likely than those who tested negative to return to the VCT site to learn their results.

¹¹ Testing and counseling in the home allows one to avoid being seen going to a clinic for a test and insures against misuse of client records by staff at clinics. On the other hand, it makes it harder to test without knowledge of one’s partner or other family members.

¹² With regard specifically to monetary aspects of cost, the contingent valuation study from Kenya of Forsythe et. al. (2002) provides evidence (though only for current VCT clients) of significant sensitivity to price.

¹³ Though the paper by Fylkesnes (2004) described above is a randomized study that provides information on preferred mode of service delivery among those who are interested in testing.

¹⁴ Though surveys can and should attempt to make some relevant unobservables 'observable', for example, surveys can collect information on perceptions on individual HIV risk.

¹⁵ Interactions among program participants that influence outcomes violate the Stable Unit Treatment Value Assumption (SUTVA) described by Rubin (1990), making causal inferences invalid in individual level studies.

¹⁶ There are other important policy questions that could in principle be addressed through community level experiments but would likely prove very costly. One concerns the appropriate role of VCT in the context of expansion of other programs such as public HIV/AIDS mass media campaigns. This as well as general secular changes in attitudes toward risky sexual behavior may mean that the marginal impacts on behavior of VCT (hence also the social return to public investments in VCT services) will fall, to the extent that behavior change is being achieved already by these other means.

¹⁷ Exceptions are likely. Some apparently healthy individuals would have an incentive to test if there is program leakage such that some who test positive are able illicitly to pay for access to subsidized drugs and begin therapy early. Also, the program will attract some individuals who believe they have AIDS related illnesses but turn out not to be ill with AIDS.

¹⁸ Consistent with this idea, surveys of commuters suggested that residents of Khayelitsha in Cape Town, where ARVS have been introduced, had more positive attitudes toward testing than in other areas throughout South Africa (see CADRE 2002).