UW Biostatistics Working Paper Series

3-18-2003

Design of the HIV Prevention Trials Network (HPTN) Protocol 054: A cluster randomized crossover trial to evaluate combined access to Nevirapine in developing countries

Jim Hughes
University of Washington, jphughes@u.washington.edu

Robert L. Goldenberg
University of Alabama, Robert.Goldenberg@ccc.uab.edu

Catherine M. Wilfert
Elizabeth Glaser Pediatric AIDS Foundation/Duke University, wilfe001@mc.duke.edu

Megan Valentine
Family Health International, mvalentine@fhi.org

Kasonde G. Mwinga
University Teaching Hospital, Lusaka, Zambia, kgmwinga@zamnet.zm

See next page for additional authors

Suggested Citation

Authors

Jim Hughes, Robert L. Goldenberg, Catherine M. Wilfert, Megan Valentine, Kasonde G. Mwinga, Laura A. Guay, Francis Mmiro, and Jeffrey S. A. Stringer

Introduction

 UNAIDS has estimated that 90% of global mother-to-child HIV transmissions (MTCT) occur in sub-Saharan Africa, with nearly 600,000 new perinatal infections annually. Essentially all of these HIV-infected infants will die prematurely, and the impact of pediatric HIV/AIDS has reversed hard-won gains in childhood mortality in much of the developing world. However, a single dose of nevirapine (NVP) given to an HIV-infected woman at the onset of labor, followed by a single dose to her infant within 72 hours of delivery can reduce MTCT of HIV by 48% relative to maternal short course zidovudine [1]. This simple and inexpensive approach to reducing MTCT is now recommended for settings in which continuous maternal antiretroviral therapy is not available [2]. However, despite the increasing availability of nevirapine or other antiretroviral interventions to prevent MTCT, many women receiving prenatal care in resource-limited settings, such as sub-Saharan Africa, do not access these therapies because they do not wish to learn their HIV status. Indeed, of 658 women offered voluntary HIV testing and counseling (VCT) in a recent study of alternate ways to administer nevirapine, only 406 (62%) chose to be tested and access nevirapine [3]. Other sites have seen even lower rates of testing and nevirapine use [4]. Thus, the standard model of requiring HIV serodiagnosis prior to initiating antiretroviral prophylaxis prevents a proven effective intervention from reaching a large proportion of HIV-infected women receiving prenatal care.

In December, 2001, the WHO convened a conference on the use of NVP for preventing MTCT of HIV in high-prevalence, low resource settings. They identified three strategies for administration of NVP: 1) targeted therapy in which NVP is offered only to women identified as HIV-seropositive through VCT; 2) universal (sometimes called "mass") therapy in which NVP is offered to all pregnant women without an option for testing; and 3) combined therapy in which VCT is made available and NVP is offered both to women who accept VCT and test positive, as well as to those women who refuse VCT. While it may seem obvious that these latter two strategies would result in the distribution of NVP to more HIV-seropositive women compared to targeted therapy, a closer look reveals some unexpected complexities. First, since NVP is self-administered by a pregnant mother at the onset of labor (a dose-timing characteristic that may be essential to its full efficacy [5]) the issue of drug adherence is important. There is some evidence to suggest that women who do not learn their status through VCT may be less likely to actually ingest the NVP tablet than those who do learn their status and thus perceive tangible infant

benefit with adherence. Second, many believe that VCT is so essential to a program of antenatal care that any strategy that might undermine its acceptability in the community should not be considered [6]. This line of reasoning would obviously disqualify the universal strategy in the minds of some decision makers. However, it is possible that offering access to NVP in a "combined " fashion, where women need not be tested for HIV, could also serve to undermine the acceptance of VCT in the community. Thus, it is unclear which strategy will result in the highest rate of NVP use among HIV-seropositive women.

HPTN054 is a randomized trial designed to compare the targeted strategy with the combined strategy for providing NVP and standardized VCT to HIV-seropositive women and their infants. The primary outcome is the proportion of HIV-seropositive women and their infants in the population who accept and adhere to the use of NVP. We refer to this proportion as "nevirapine coverage". Nevirapine coverage will be assessed by collecting anonymous, unlinked cord blood specimens from all women receiving care and delivering at participating clinics. Positive maternal HIV status (the coverage denominator) as well as presence of NVP (the coverage numerator) will be obtained from these cord blood specimens. Directly observed therapy will be used to assess infant receipt of NVP. Uptake of VCT will also be assessed as a secondary endpoint. A number of unique issues have arisen during the design of this trial. In the following sections we discuss each of these in turn.

Need for a cluster-randomized trial

A cluster randomized trial is generally less efficient, more complex and more expensive than a comparable individually randomized trial [7]. Cluster randomized trials are typically best suited for situations where randomization of individuals is not feasible either for logistical reasons or due to fear of "contamination" of the treatment effect [8]. Our setting meets both of these criteria. In the typical sub-Saharan African prenatal clinic, women receive most health care education and counseling in groups (including offers for HIV pre-test counseling). Thus, to individualize the message would not reflect the normal conduct of prenatal care in this setting. In addition, women may come to the clinic for more than one prenatal care visit and it would be difficult in this setting to keep track of their randomization status from visit to visit. We also know from experience that women commonly discuss their care among themselves while waiting to see the clinician. Contamination could easily occur if women who were randomized to the targeted arm and chose not to be tested asked friends in the combined arm to obtain NVP for them. Thus, we

1

chose a cluster-randomized design in which the basic unit of randomization is the prenatal care clinic. Participating facilities have been selected for geographical and/or social isolation from other clinics (to minimize treatment selection by the participants and migration among clinics), high anticipated rates of participation (see below) and high anticipated retention through delivery.

Need for a population-based study with high levels of participation

Most clinical trials ensure internal validity via randomization but are less concerned about external validity. Indeed, the self-selection that follows from the enrollment and consent process usually ensures that subjects participating in the study are different in a number of ways from non-participants. Often this does not make much difference if the treatment effect is expected to be similar among those that chose to participate and those that refuse. In this study, however, a key feature of the combined strategy is that women who refuse HIV testing can still receive NVP. We believe that women who refuse testing are likely the same women who would be most likely to refuse to participate in a research study, particularly if the study would result in disclosure of the woman's HIV status. Further, the aim of HPTN054 is to measure the expected nevirapine coverage rates that would be expected in practice, not in the context of a research study. Thus, even moderate refusal rates could lead to biased estimates of the treatment effect in spite of randomization, as shown in table 1. This table illustrates how incorrect conclusions could be drawn about the population coverage of the two NVP strategies if we were to consider only those women who are interested in participating in a research protocol. Similar rates of nevirapine coverage are observed in the two treatment arms among women who are (theoretically) willing to participate in a research study. However, among women who would (theoretically) refuse to participate in a research study, lower rates of coverage are seen under the targeted arm compared to the combined arm since these are likely the same women who would refuse VCT. The net result is that, in the total population, coverage rates are substantially higher in the combined arm even though no difference is seen in the subpopulation of women willing to participate in a research study.

As a result of these considerations, HPTN054 has been designed to achieve very high participation rates. The most important strategy being used is that women can choose to participate in part, but not all, of the study. In particular, all women attending a study clinic for prenatal care will be asked to consent to provide an anonymous, unlinked cord blood sample at delivery. Collection of these samples will provide the outcome data for the primary aim of this

study (see further discussion below) without requiring the women to be tested and become aware of their HIV status. We expect very high rates of participation in this phase of the study. Following delivery, women will be offered full participation in the study, which will involve linkage of demographic data, testing results and followup. We expect rates of participation in this second phase to be more typical of research projects in this region (~50%).

Table 1. Illustrative example showing the bias in the comparison of treatments that can result if there is a significant interaction between willingness to participate in a research study and willingness to be tested for HIV.					
	Willing to participate (N = 60)			Not willing to participate (N = 40)	
	NVP +	NVP -		NVP +	NVP -
Targeted	22	8		1	19
Combined	22	8		13	7

Population Definition

When the primary outcome in a cluster randomized design is measured on individuals within the cluster, it is important to have a careful definition of cluster membership. In HPTN054 we need to uniquely define the population of women who attend a given clinic during a particular time period. We have chosen to use the following definition: a woman is a member of the population of a given clinic during a given time period if she presents for her first antenatal visit at that clinic during that time period. An alternative definition would be to include all women who attended any antenatal visit during the time period at the clinic. However, since women are counseled and offered HIV testing at each prenatal visit, the latter approach could give a distorted picture of the operating characteristics of the interventions. That is, women who are near to delivery at the start of the study would receive fewer exposures to the intervention message than would occur in routine practice. These women may be less likely to accept testing and/or adhere to NVP usage. Women who present for their first antenatal visit at one clinic but then attend another clinic for subsequent visits will still be considered as members of the population of the clinic where they first received care. Although this definition potentially allows women to select their treatment based on their choice of antenatal clinics, we will attempt to minimize this possibility by choosing clinics that are geographically and socially distant from each other. Since women typically attend their first prenatal visit around 27 – 28 weeks of gestation, an important consequence of this

3

population definition is that the intervention must continue for at least 3 months following the enrollment period to ensure that women receive a consistent counseling message during their entire pregnancy.

Collection of endpoint information

As noted above we have divided the study into two parts to ensure high levels of participation for the primary endpoint. The first part consists of collection of an anonymous, unlinked cord blood sample from virtually all women delivering at the clinic. These cord blood samples provide three of the four key measurements required for this study: treatment arm (by knowing which clinic the cord blood sample came from), maternal HIV status and maternal NVP uptake. The last key measurement is infant NVP uptake, which will be ascertained by direct observation. To link infant NVP uptake to the maternal cord blood sample we will attach identical numbered stickers to the cord blood sample and the maternal/infant medical chart at the time of delivery. When the infant is given his or her dose of NVP (typically within 24 hours) the numbered sticker will be removed from the medical chart and attached to the cord blood sample. Thus, the presence of the second sticker on the cord blood sample will be evidence that the infant dose was given. At the same time this procedure breaks the link between the cord blood sample and the maternal chart ensuring that the sample is unlinked and anonymous. No other record will be kept linking the cord blood sample with the woman who provided the sample.

The primary endpoint in this trial – presence of nevirapine in the mother and infant – is clearly a surrogate for the clinical endpoint of interest, HIV infection in the infant. However, previous studies have clearly demonstrated the efficacy of single dose nevirapine for reducing MTCT of HIV. In addition, use of a surrogate in this trial is more than merely a matter of convenience. Ascertainment of HIV infection in the infant would require followup of the mother infant pair with attendant loss of anonymity. This would almost certainly lead to lower participation rates in the trial, which, as we argued in table 1, might bias the results. Thus, the use of anonymous, unlinked cord blood samples to obtain the surrogate endpoint of nevirapine coverage is a key component of the design.

In the second portion of the study, women will be asked to participate in a linked study that will provide more detailed information and followup for safety endpoints. We expect participation in this portion of the study to be lower since it requires identification and followup. By separating

4

the two portions of the study and having separate consent forms for each, we hope to have very high participation rates for the primary endpoint - nevirapine coverage among HIV-seropositive women.

Statistical Design

In this section we describe three possible designs for this trial and outline the advantages and disadvantages of each.

Parallel - The term parallel design is used to refer to the standard setting in which 2N clusters or communities are enrolled in the study and half (N) are randomized to treatment A and half are randomized to treatment B. In a matched parallel design, the communities are first matched and randomization is done within the matched sets.

Crossover - In a crossover design, each cluster receives both interventions in a random order. Thus only N clusters are required for this design. Each cluster serves as its own control and in this sense a crossover design is similar to a matched-pairs parallel design. However, since each cluster must provide both interventions, the duration of a study using a crossover design will typically be longer than a study using a parallel design. Also, issues of time trends and washout effects must be considered in a crossover design (see below). One can envision a crossover design in which the individuals receiving the intervention and providing outcome information are the same under both treatments, or, as in the case of HPTN054, change between treatment periods.

Stepped wedge – The term stepped wedge is used to refer to crossover designs in which the crossovers occur in one direction only (i.e. from A to B but not B to A). However, the time of the crossover is randomized. Issues of time trends and washout effects must be considered in this design as well.

These three designs are illustrated schematically in figure 1.

The following basic model may be used to describe Y_{ij}, the cluster level response for cluster i at time j, for designs of the form shown in figure 1

$$Y_{ij} = \mu + \alpha_i + \beta_j + \theta X_{ij} + e_{ij} \tag{1}$$

where $\alpha_i \sim N(0,\tau^2)$ is a random clinic effect (i = 1 … N), β_j is an arbitrary time effect (j = 1 … T) and θ is the treatment effect. The $e_{ij} \sim N(0,\sigma^2)$ are independent random errors. X_{ij} is a treatment indicator such that X_{ij} is 0 if cluster i is given treatment A at time j and 1 if cluster i is given treatment B at time j. That is, X describes the change in the treatments for each site over time. τ^2 is often referred to as the "between-cluster" variation while σ^2 is the "within-cluster" variation. Sample size calculations for cluster randomized trials are often complicated by the requirement that good estimates of both of these sources of variation must be available a priori. This issue is discussed further below.

Figure 1. Alternative designs. A and B represent the two randomization arms.

		Parallel			Crossover				Stepped wedge			
		Time			Time				Time			
		1			1	2			1	2	3	4
	1	A		1	A	B		1	A	B	B	B
Cluster	2	A		2	B	A		2	A	A	B	B
	3	B		3	A	B		3	A	A	A	B
	4	B		4	B	A						

A parallel design is the most straightforward approach for many trials. Nonetheless, sample size determination for parallel trials is challenging since the required number of sites is highly dependent on the magnitude of the between site variance, τ^2 (equivalently, the intraclass correlation, $\tau^2/(\tau^2 + \sigma^2)$, or between site coefficient of variation, τ/μ). Matching can sometimes be used to reduce this dependence but the degree of reduction depends on the efficacy of the matching. A major advantage of the parallel design is that the final analysis of the (cluster-level) outcomes is a simple t-test.

The sample size for a crossover design does not depend on the between-site variance since the treatment effect is estimated from within-site comparisons only. However, a crossover trial is not always feasible and will, in any event, take at least twice as long to conduct as a comparable parallel trial. The analysis of the (cluster-level) outcome data from a crossover trial is straightforward (a simple t-test on the within-cluster differences between treatments) provided there are no time trends and/or carryover effects. However, in the presence of a time trend the

expected value of the within site differences between the treatment arms will differ depending on the order of the treatments. Although the simple t-test analysis remains unbiased, it is inefficient and a more complex analysis that involves modeling the time effect is required to gain efficiency. Carryover effects occur when the effect of the first treatment persists into the second intervention period. In terms of the model (1) such effects would be manifest as a treatment by time interaction term and, again, a relatively complex analysis is required. Often it is possible to eliminate potential carryover effects in the design phase by including a washout period between the two intervention periods rather than by trying to control for the carryover effect in the analysis phase. For instance, in HPTN054 a washout period of approximately 3 months is required after the end of the first intervention period to allow women enrolled during this period to deliver their infants. The first intervention would be continued during the washout period to ensure that these women received a consistent counseling message. However, women who initiated prenatal care during the washout period would not be considered part of the study population.

Stepped wedge designs are not commonly used but may arise from certain practical considerations. For instance, in a hepatitis B virus (HBV) vaccine intervention trial in Gambia [9], the "A" treatment corresponded to the existing standard of care (no vaccine) while the "B" treatment was the HBV vaccine. However, it was not possible to implement the new vaccination program over the whole of Gambia concurrently. Rather, the HBV vaccine was phased in over 4 years in 17 health districts and the timing of the vaccine introduction was randomized among the 17 districts. Stepped wedge designs are less dependent on between site variation than parallel designs. However, carryover effects and time trends are a concern with these designs as well. Although a simple within-cluster analysis is possible in the absence of a time trend, such an analysis will be biased if a time trend is present. In most cases, therefore, a more complex analysis based on the model (1) must be used. Finally, since at least two crossover times (i.e. 3 time periods) are necessary to ensure that the treatment effect is not confounded with an underlying time trend, trial duration may be a concern with stepped wedge designs as well.

Power and Sample Size

Typically, the aim of a trial is to test for a significant treatment effect. This is equivalent to testing the hypothesis $H_o: \theta = 0$ versus $H_a: \theta = \theta_a$ in model (1). Under model (1) it can be shown that the variance of the estimated treatment effect, $\hat{\theta}$, is

$$\text{Var}(\hat{\theta}) = \frac{N\sigma^2(\sigma^2 + T\tau^2)}{(NV - W)\sigma^2 + (V^2 + NTV - TW - NU)\tau^2} \qquad (2)$$

where N is the number of clinics, T is the number of time periods, and $V = \sum_{ij} X_{ij}$,

$W = \sum_j (\sum_i X_{ij})^2$, and $U = \sum_i (\sum_j X_{ij})^2$. Given a particular design (represented by X) and

hypothesized values of σ^2 and τ^2, the approximate power to test the hypothesis $H_o: \theta = 0$ is (for a

two-tailed α-level test)

$$\text{power} = \Phi^{-1}\left(\sqrt{\frac{\theta_a^2}{\text{Var}(\hat{\theta})}} - Z_{1-\alpha/2}\right) \qquad (3)$$

where Φ is the normal distribution function and Z_p is the p'th quantile of the normal distribution

function. Use of normal quantiles rather than the quantiles of the t-distribution in (3) can result in

an overestimate of the power when N, the number of clinics, is small. Snedecor and Cochran [10]

recommend substituting N' = N – 1 for N in the variance formula in this case.

Example: HPTN054

In the case of the HPTN054 trial, little data on the expected interclinic variation in NVP coverage

were available when the study was being designed. The data that were available consisted of

uptake of VCT among women of unknown serostatus from prenatal care clinics in a number of

countries. Not only is this a different outcome on a different population from the HPTN054 trial,

but counseling practices varied significantly from clinic to clinic and country to country. An

analysis of these data gave, not surprisingly, a very large estimate of clinic to clinic variation

(specifically, the coefficient of variation in VCT uptake was estimated as 50% using the

estimation procedure in Hayes and Bennett [11]). Although it was felt that the variation would be

less in HPTN054, where a standardized counseling message in a more homogeneous setting

would be used, the extent of the expected reduction was unknown. Further, sample size

calculations showed that even a moderate degree of inter-clinic variation would necessitate a

large number of clinics in a parallel design and this was not feasible since only a limited number

of clinics with the desired patient population characteristics (geographically or socially isolated,

high expected participation rates, high retention through delivery) were available. A crossover

8

design was considered since this would require fewer clinics and would not require a prior estimate of inter-clinic variation. However, community representatives warned that while crossing over from the targeted arm to the combined arm was acceptable, crossing over from the combined arm to the targeted arm would be unacceptable in the community, since it would appear that a previously available option (availability of NVP without being tested for HIV) was being withheld. Starting the study with the combined approach was also a concern among some health care providers whose current standard of care promoted the targeted approach.

The above issues led to the consideration of a stepped wedge design in which crossovers would occur in only one direction. However, since each crossover period requires a minimum 3-month washout (as described above), we wanted to minimize the number of crossovers. The final design for this study is a combination of the parallel and stepped wedge designs and will be conducted in Lusaka, Zambia and Kampala, Uganda (figure 2). The basic rationale for this design is that the presence of the parallel design clinics (1 and 4) allows us to control for underlying time trends while the crossover clinics (2 and 3) provide more precise within-clinic comparisons between the treatment arms.

Figure 2. Design of HPTN 054. T = Targeted therapy; C = Combined therapy

		Time	
Site	Clinic	1	2
	1	T	T
	2	T	C
Kampala	3	T	C
	4	C	C
	1	T	T
Lusaka	2	T	C
	3	T	C
	4	C	C

Using equation (2), the variance of the estimated treatment effect for this design is

9

$$\text{Var}(\hat{\theta}) = \frac{\sigma^2(2\tau^2 + \sigma^2)}{2\tau^2 + 3\sigma^2}$$

Since the result for each clinic-time period is a sample proportion, we have $\sigma^2 = p(1-p)/m$ where m is the number of HIV-seropositive women per clinic-time period and p is the nevirapine coverage rate. We expect p in the range 0.4 - 0.7 so it is conservative to substitute p = .5 in this formula. If we parameterize the between-clinic variation in terms of the coefficient of variation, k = τ/p, then the above expression may be written as

$$\text{Var}(\hat{\theta}) = \frac{2mk^2 + 1}{4m(2mk^2 + 3)}$$

Our goal is to enroll 38 HIV-seropositive women per clinic-time period which gives at least 90% power across a wide range of values for k against the alternative hypothesis H_a: $\theta = 0.25$ (figure 3). For comparative purposes, this figure also shows the power for a 16 clinic, single time period parallel design. The parallel design has slightly higher power for k < .23 but has dramatically lower power as k increases. This occurs because the within-clinic variation (σ^2) dominates the variance for k < 0.2 while the between clinic variance (τ^2) dominates for larger k. Note that the total number of subjects (16*38 = 608) is the same for both designs.
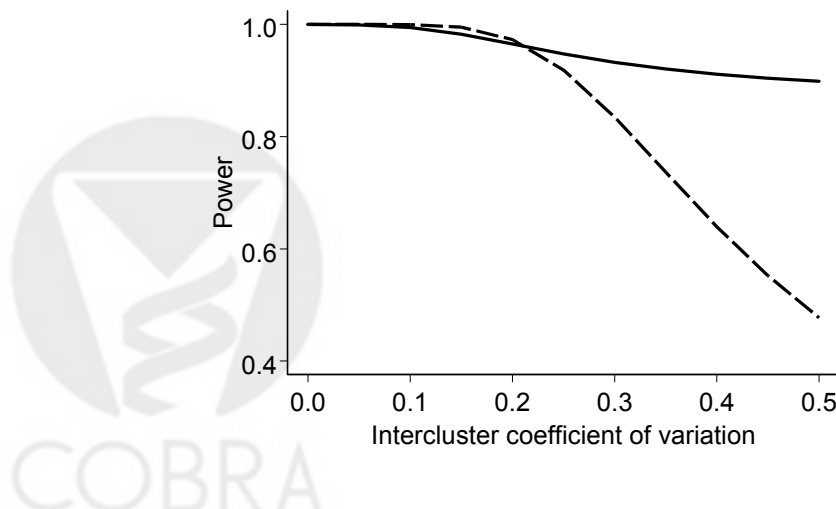

Figure 3. Power of the HPTN054 study design shown in figure 2 against the alternative hypothesis H_a: $\theta = .25$ (solid line). Also, shown is the power for a comparable 16 clinic parallel design (dashed line). In each case we assume m = 38 HIV-seropositive women per clinic per time period.

A number of assumptions underlie the design for HPTN054. Most importantly, we assume that the model (1) can be used to describe the clinic-level responses. For the purposes of study planning, this model has been kept relatively simple. However, in the analysis phase it will be possible to extend this model to check for a site effect, time by site interaction and time by treatment interaction (although such tests have relatively low power). We assume there are no time by clinic interactions (i.e. the time trend, if it exists, is similar at all clinics, at least within a site). To avoid a carryover effect a 3-month washout period will be included following time period 1 before starting the intervention in time period 2. The washout period should allow virtually all of the women enrolled in time period 1 to deliver their infants before the second intervention period begins.

Discussion

Although cluster randomization is becoming more and more common, particularly in disease prevention settings, the use of any type of crossover design in cluster randomized trials is unusual. Typically, crossovers are either not scientifically defensible (e.g. because more susceptible subjects are removed from the population during the first time period, a type of carryover effect) or not feasible because they would take too long (e.g. when a lengthy intervention and followup period are necessary). When a crossover trial is feasible, however, designs of this type have some distinct advantages. Often, there is a high cost associated with adding additional clinics to a study but a relatively low cost associated with enrolling more subjects at an established clinic. Thus, crossover designs can be more cost efficient than parallel designs. In addition, as we have noted, crossover trials are less sensitive to between clinic variation. Matched parallel designs have this same feature provided appropriate matching criteria can be identified *a priori*. This may not be easy. In the case of HPTN054, a limited number of clinics, inadequate prior information on between clinic variation and community resistance to crossover from the combined to targeted arm all contributed to the mixed parallel/stepped wedge design shown in figure 2. Just as importantly, however, high HIV prevalence at these sites means that adequate enrollment of HIV-seropositive women can be achieved in only 4 months in each intervention period. If we add 3 months after each intervention period to allow women to deliver, the total study duration will still be only about 14 months. Thus, the mixed parallel/stepped wedge design uses fewer resources, requires less information a priori and still provides a well-powered study that can be completed in a relatively brief period.

11

References:

1. Guay LA, Musoke P, Fleming T, Bagenda D, Allen M, Nakabiito C, et al., 1999. Intrapartum and neonatal single-dose nevirapine compared with zidovudine for prevention of mother-to-child transmission of HIV-1 in Kampala, Uganda: HIVNET 012 randomised trial. *Lancet* 354:795-802.

2. World Health Organization, 2001. Consultative meeting on the use of nevirapine for the prevention of mother-to-child transmission of HIV among women of unknown serostatus, draft report. Geneva, Switzerland December 5-6

3. Stringer JSA, Sinkala M, Goldenberg RL, Stout JP, Kumwenda R, Mwinga K, Aldrovandi G, Vermund SH, 2001a. A pilot study of nevirapine administered upon presentation in labor without HIV testing. 3rd Conference on Global Strategies for the Prevention of HIV Transmission from Mothers to Infants; Kampala, Uganda. September 9-13. *Abstract 300*.

4. Bassett MT, 2002. Ensuring a public health impact of programs to reduce HIV transmission from mothers to infants: The place of voluntary counseling and testing. *Amer J Public Health* 92:347-351.

5. Stringer JSA, Sinkala M, Goldenberg RL, Stout JP, Acosta EP, Kumwenda R, Cliver S, Aldrovandi GM, Mwinga K, Vermund SH, 2001. A prospective comparison of two strategies for perinatal nevirapine administration in Zambia. 3rd Conference on Global Strategies for the Prevention of HIV Transmission from Mothers to Infants; Kampala, Uganda. September 9-13. *Abstract 303*.

6. World Health Organization, 2001. New data on the prevention of mother-to-child transmission of HIV and their policy implications. Conclusions and recommendations. WHO technical consultation on behalf of the UNFPA/UNICEF/WHO/UNAIDS Inter-Agency Task Team on Mother-to-Child Transmission of HIV. Geneva, 11-13 October 2000. Final report, WHO Geneva, January.

7. Donner A, Klar N, 2000. Design and Analysis of Cluster Randomized Trials in Health Research, Arnold Publishers Limited, London.

8. Torgerson DJ, 2001. Contamination in trials: Is cluster randomization the answer? *British J Medicine* 322:355-357.

9. Gambia Hepatitis Study Group, 1987. The Gambia Hepatitis Intervention Study. *Cancer Research* 47:5782 – 5787.

10. Snedecor G, Cochran W, 1989. Statistical Methods (8th ed), Iowa State University Press, Ames, Iowa.

11. Hayes RJ, Bennett S, 1999. Simple sample size calculations for cluster-randomized trials. *Int. J Epidemiology* 28:319-326.