**Supplemental Digital Content 1**

**1.1 Population-Related Study Parameters**

HIV Risk Groups

HIV risk groups are defined as subsets of the study population with similar underlying risk of becoming infected relative to the rest of the population. The term “risk groups” should not be confused with the term “transmission categories” such as men who have sex with men (MSM), injecting drug users (IDU), etc. For the purpose of this analysis, the term “risk group” is used to define a subset of the population with similar probability of disease regardless of behavior. We define these subsets as having HIV risk ranging from very low to very high. If the number of risk groups is r, risk groups are identified using the indexing variable, *i* ,  *i=1,…r*. Follow-up time intervals are identified using the indexing variable, *t*. In order to model the change in the risk distribution of a hypothetical study population over time, our mathematical formulation includes the proportion of participants () in each study arm belonging to the *i*th risk group. In the presence of frailty, these proportions change over time (See Figure 4). The proportionof participants in a study arm belonging to the *i*th risk group at time, t, is, therefore, denoted .

Probability of Disease for risk group *i* at time *t*

At time *t*, the probability of disease for the *ith* risk group () is defined as the probability that an individual within the *ith* risk group becomes infected after time t but prior to time t+1 (the symbol “|” indicates that the probability is conditional on or depends on time). The probability of HIV infection within a population can vary widely depending on demographic factors such as gender, sexual and drug use behavior, and geographic region (a marker for regional norms, healthcare access, HIV prevalence, etc.) [1]. Much of the current literature on HIV transmission probability focuses on estimating the probability of HIV transmission per risky act [1, 2]. Here, *πi|t*  is interpreted as the average probability of HIV infection for all risky acts performed by all subjects within a risk group after time *t* but prior to time t+1 As a result, we assume that at any time *t*, all subjects within a risk group have the same probability of disease.

Frailty

Frailty is defined as the degree of heterogeneity in the probability of disease among the risk groups. In order to incorporate different degrees of frailty, we constructed a series of risk profiles that reflect varying degrees of heterogeneity in infection risk. These profiles were constructed by first selecting a disease probability for the highest risk group. We then calculated the infection probabilities for all subsequent risk groups by sequentially multiplying by a fixed multiplier, 1- *η*, thereby creating a geometric progression of probabilities ranging from slightly to dramatically lower disease probability relative to the disease probability of the highest risk group. For example, if the disease probability of the highest risk group is .60 and *η* =0.2, the probability of disease for the second highest risk group is 20% lower than that of the highest risk group (or 48%). The third highest risk group has a probability of disease that is 20% lower than that of the second (or 38%), and so on. Constructing the risk distributions in this way (1) establishes a fixed proportion by which adjacent risk groups differ in their probability of disease, (2) creates a means of varying risk heterogeneity (frailty) through a single parameter, *η*, which is directly related to frailty (i.e., a higher *η* corresponds to greater frailty), and (3) represents absence of frailty when *η* equals zero.

**1.1 Intervention-Related Parameters**

Intervention efficacy

Intervention efficacy is the percent reduction in the probability of disease conferred by the intervention. In the context of our defined study parameters, if the intervention reduces the probability of disease by a factor of ε, the new disease probability can be calculated by multiplying the original probability by 1- ε. For example, if ε = 80%, then the probability of disease for individuals in the *i*th risk group at *t* \* (1-.80) or (\*.20.

Waning (*ω*):

Waning is defined as the reduction in the intervention efficacy (which translates into a corresponding increase in the probability of disease) that occurs in the intervention arm over time. It is assumed that the underlying *ω* is constant across risk groups. In other words, the intervention is equally effective at preventing disease regardless of individual disease risk. When waning is present, intervention efficacy at each time, *t,* is computed as the product of the efficacy at time 0 and the proportion of intervention efficacy remaining at time t

(1)

where is the efficacy of the intervention at time *t* and is its efficacy at t=0. For example, if intervention efficacy at time 0, , is 50% and this efficacy wanes at a rate of 10% per time interval, (i.e., ω=.10), the efficacy of the intervention at say t=5, , is the initial intervention efficacy multiplied by a factor of So, rather than being 50% as it was at time 0, the intervention efficacy at t=5 is . In other words, for those subjects in the treatment arm who remain at risk of HIV infection at t=5, their probability of infection is only reduced by ~30% because of the intervention.

**1.2 Outcome Measures**

1.2.1 Risk Group-Specific Disease Incidence (*Ii*)

To assess the impact of frailty on disease incidence over time, we first assume that all participants are disease free at the start of the study. We then calculate the risk group-specific disease probabilities at time 0 (i.e., the probability that a member of risk group becomes infected prior to time t=1) for both treatment arms. The risk group-specific disease probability at t=0 is the product of the disease probability of the highest risk group, , and the frailty multiplier, , corresponding to that risk group

(2)

In the placebo arm, for subsequent follow-up times, the risk group-specific disease probability is computed using the same formula. For the intervention arm, we also incorporate an adjustment for the intervention efficacy. If, at time *t*, the treatment prevents infection with an efficacy of εt,, then the probability of disease in the intervention arm is reduced by a factor of 1-εt.

(3)

To allow the proportion of subjects in each risk group to vary, we must weight each probability by the proportion of the at-risk population belonging to the risk group at time *t*. For our purposes, this proportion is denoted and for the placebo and intervention groups, respectively. The overall time-specific incidence rate is computed as the sum of these values across risk groups. We refer to the time-specific incidence rates for the placebo and intervention arm as , respectively.

(4)

(5)

1.2.2 Time-specific Incidence Rate Ratio (*IRR*)

The IRR at time *t* comparing the intervention and placebo arm is the ratio of equations (5) and (4).

= (6)

This formulation of the IRR has two noteworthy properties. First, it is independent of sample size. Second, it is independent of the value of the disease probability in the highest risk group as this value cancels out from the numerator and denominator. In the absence of frailty (i.e., probability of infection is constant across all risk groups, =0), the IRR reduces to . To see this, note that in this case, the risk groups have the same probability of infection, so the population reduces to a single risk group and the summation sign is no longer needed. In addition, the proportion of the population belonging to this single risk group in each arm, and , are both 100%. The resulting IRR then reduces to . In the absence of both waning and frailty, the IRR reduces to, the commonly accepted definition of the IRR. Using this formulation, our IRR is a function of the degree of frailty, η, and the efficacy of the intervention, εt, which is inversely proportional to waning.

References:1Boily, M.C., et al., Heterosexual risk of HIV-1 infection per sexual act: systematic review and meta-analysis of observational studies. *Lancet Infect Dis*, 2009; 9(2): 118-29.

2Baggaley, R.F., White, R.G., and Boily, M.C. HIV transmission risk through anal intercourse: systematic review, meta-analysis and implications for HIV prevention. *Int J Epidemiol*, 2010; 39(4):1048-63.

**Supplemental Digital Content 2**

The following are screenshots of the Microsoft Excel 2010 application used to compute intermediate probabilities as well as perform final incidence calculations at baseline (or Time 0) and the first three follow-up times. (The actual analysis includes 10 follow-up times). The underlying calculations were performed using macros written in Visual Basic for Applications. Using these macros, we were able to automate the process of manipulating, analyzing and outputting the values obtained in each study scenario.

**Figure 1a. “No Frailty/No Waning” Scenario**



\*The weighted disease probability is the contribution that each risk group makes to the total disease incidence at time t. This intermediate calculation is the product of the proportion of the population belonging to that risk group and the corresponding probability of disease at time t. The incidence at time t is the sum of these values across risk groups (see Equation 4).

**Figure 1b. “Moderate Frailty/30% Waning” Scenario**



\*The weighted disease probability is the contribution that each risk group makes to the total disease incidence at time t. This intermediate calculation is the product of the proportion of the population belonging to that risk group and the corresponding probability of disease at time t. The incidence at time t is the sum of these values across risk groups (see Equation 4).

Figure 1a contains intermediate and final calculations under the “No Frailty/No Waning” scenario. Figure 1b contains the calculations for the “Moderate Frailty/30% Waning”. The macro input values are located in the box at the upper-left of the spreadsheet. The application allows one to vary all of the input values but, for the purposes of this analysis, we held certain study parameters fixed. The fixed study parameters include: intervention efficacy (ε=0.5) , time interval length (1 year) and the probability of infection for the highest risk group ( The user-specified study parameters include: waning (*ω*) and frailty (*η*).

No Frailty/No Waning

The top and bottom sets of rows show the calculations for the placebo and treatment group, respectively. The first column enumerates the separate rows for each of the 5 risk groups (Assumption #1). The next 3 columns outline the risk distribution, probability of disease, and intervention efficacy at baseline (Assumptions #2, #3 and #5). Note: for the placebo group, the column for intervention efficacy is left blank because this group by definition does not receive an intervention. The remaining columns show the intermediate calculations involved in computing the disease incidence at each follow up time. In the absence of both frailty and waning, incidence and the resulting IRR remain constant across all time points.

Moderate Frailty/30% Waning

Figure 1b differs from Figure 1a in that these calculations incorporate both a frailty and waning component (see input values for variable study parameters). With the inclusion of both risk heterogeneity (frailty) and waning intervention efficacy, the incidence in the treatment and placebo groups get closer and the IRR approaches 1.

**Supplemental Digital Content 3**

Using the disease probabilities and IRRs from the various study scenarios, we implemented the power calculation method described by Rosner [1]. We wish to test the hypothesis H0: RR=1 versus H1: RR≠1 (two-sided alternative) for a specified significance level, α=.05. Assuming an intervention trial with 5-year follow up, the statistical power of a test is:

Power= *z*1-α/2)

Where k is the ratio of participants in the intervention group compared to the placebo group and m is the total number of events (or number of subjects who become infected) during the study. IRR(t) is the incidence rate ratio comparing the intervention to the placebo group at time t. is the cumulative distribution function for the standard normal distribution. (Note: All of our scenarios assume equal allocation so, for our purposes, k is always equal to 1.)

Remaining consistent with our prior notation, the number of subjects that become infected at time, t, for the intervention and placebo group, respectively are:

mI(t)= nI(t) \* II(t)

mP(t)= np(t) \* Ip(t)

where nI(t) equals the number of subjects at risk at time t for the intervention group and nP(t) represents the same for the placebo group. Consequently, the total number of subjects who become infected during the study, m, equals ∑ mI(t)+ mP(t), across of values of t.

1Rosner B. (2006). Fundamentals of Biostatistics. (6-th edition). Thomson Brooks/Cole.