

# How to Pay “Cash-on-Delivery” for HIV Infections Averted: Two Measurement Approaches and Ten Payout Functions

**Timothy B. Hallett and Mead Over**

## Abstract

In contrast to current donor policy, which funds a recipient country’s national AIDS control program, this paper proposes a measurement strategy to enable a donor to reward a recipient country’s success at HIV prevention, irrespective of the inputs, activities, or who gets the credit. In accordance with the “cash-on-delivery” model of foreign assistance, the objective is not to replace traditional input- or activity-oriented aid, but to complement it by enhancing the motivation for local actors and their partners (including the traditional bilateral and multilateral funding agencies and their agents) to achieve measurable reductions in the rate of new HIV infections. This paper proposes two approaches to measuring the number of HIV infections averted between a baseline survey and a follow-up survey and explores the properties of ten alternative “payout functions” which would link measured epidemic changes to the size of the reward to be paid. All measurement approaches include the possibility of statistical error and thus a risk of rewarding the country too little or too much. This risk depends on the initial rate of infection and on HIV prevention success and can be reduced by either increasing the survey sample size or increasing the interval between surveys. By negotiating in advance the choice of one of these measurement approaches and one of a menu of payout functions, the donor and recipient agree on the recipient’s incentive structure with respect to the magnitude and precision of the estimated reduction in the rate of new infection.

**How to Pay “Cash-on-Delivery” for HIV Infections Averted:  
Two Measurement Approaches and Ten Payout Functions**

Timothy B. Hallett  
Institute for Global Health  
Imperial College London

Mead Over  
Center for Global Development  
Washington, D.C.

Correspondence: [mover@cgdev.org](mailto:mover@cgdev.org)

CGD is grateful for contributions from the Bill & Melinda Gates Foundation in support of this work.

Timothy B. Hallett and Mead Over. 2010. “How to Pay “Cash-on-Delivery” for HIV Infections Averted: Two Measurement Approaches and Ten Payout Functions.” CGD Working Paper 210. Washington, D.C.: Center for Global Development.  
<http://www.cgdev.org/content/publications/detail/1424088>

**Center for Global Development  
1800 Massachusetts Ave., NW  
Washington, DC 20036**

202.416.4000  
(f) 202.416.4050

**[www.cgdev.org](http://www.cgdev.org)**

The Center for Global Development is an independent, nonprofit policy research organization dedicated to reducing global poverty and inequality and to making globalization work for the poor. Use and dissemination of this Working Paper is encouraged; however, reproduced copies may not be used for commercial purposes. Further usage is permitted under the terms of the Creative Commons License.

The views expressed in this paper are those of the author and should not be attributed to the board of directors or funders of the Center for Global Development.

### ***A Story about the Power of Economic Incentive***

*From time immemorial the inhabitants of Aneityum, an island of the South Pacific nation of Vanuatu, tolerated endemic malaria, with its periodic bouts of fever and occasional childhood deaths. In the 1980s, their island lifestyle was sustained by fishing and enriched by bi-annual visits of a cruise ship and its free-spending tourists. Then in 1990 two tourists contracted malaria. According to Kevin Palmer of the World Health Organization, "I talked to the captain who confirmed that passengers had gotten sick with malaria after visiting Vanuatu. It wasn't certain whether they were infected during the Aneityum visit or not, but regardless [the cruise ship captain] demanded that the island be declared malaria free or the ship wouldn't stop anymore. This is why the islanders came to WHO and the malaria program people in Port Vila [their country's capital city] to ask for help." [1] Journal articles document how the extraordinary community-level motivation and participation of the Aneityum population led to the successful elimination of malaria on the island. [10] According to Palmer, "The happy ending to the story is that Aneityum is now a popular tour destination known as Mystery Island so the people on Aneityum are economically secure thanks to what they accomplished." [1]*

### **Introduction**

In recent years there have been billion dollar investments in interventions to treat AIDS patients and reduce HIV transmission, including the President's Emergency Plan for AIDS Relief (PEPFAR) and the Global Fund to Fight AIDS, Tuberculosis and Malaria (GFATM) [2,3,4,5]. The promise from early modelling work [6,7,8,9] was that a scale-up of interventions and services, thought to be effective at reducing the risk of infection, would reduce transmission throughout the population. While the greatly increased donor spending has dramatically increased the number of persons in poor countries who receive antiretroviral therapy (ART), observers agree that the donors have had difficulty pointing to clear successes in HIV prevention. Since the only sure way to reduce the future cost of ART in AIDS affected countries is to prevent infection, the need for effective intervention is becoming more acute and apparent.

This paper proposes the application of a new approach to foreign assistance, called the “Cash-on-Delivery” (COD) approach, to the arena of HIV prevention. The COD approach has been developed and proposed for application in other sectors in response to four intellectual currents in the field of foreign assistance in general and HIV prevention in particular [11,12]. First, there is a trend towards results-based foreign assistance, which is exemplified by the US government’s Millennium Challenge Corporation [13] and the Global Alliance for Vaccines and Immunisation at the national level and the performance based incentive movement at the level of the individual social service facility [14]. Second, there is a growing belief among some observers that the foreign “planners” who have traditionally led the design and implementation of development assistance, have less pertinent knowledge and are less important for success than are the local “seekers,” who could best solve the problems of making schools or health clinics work on the ground, provided they were properly motivated [15]. Third, following the work of Filmer and Pritchett, an increasing number of observers have wondered whether the poor results achieved by public service delivery in poor countries can be partly attributed to insufficient motivation at higher as well as lower levels of public service delivery institutions [16]. Finally, in the domain of HIV prevention, analysis of data on the cost of service delivery reveals enormous unexplained variation in unit costs for the delivery of the same service within as well as across countries, suggesting substantial scope for enhanced incentives to improve the prevention results obtained from available prevention resources [17,18,19,20].

The cash-on-delivery approach is a natural consequence of these ideas. The basic concept is that a donor (bilateral, multilateral or philanthropic) and recipient (national or state government) enter into an agreement or “contract” in which the donor agrees to reward the recipient country if it achieves, or proportional to its achievement of, a certain development outcome. As distinct from financing budget support, inputs or activities, the COD approach to assistance pays only on the basis of measured results. The COD payment does not substitute for traditional input-based foreign assistance. Indeed the intent of COD is to improve the effectiveness of other aid modalities by motivating national leaders and local actors to focus on the HIV prevention objective in the interest of obtaining the prize. Because the COD approach does not prescribe how prevention results are to be achieved or that any reduction in incidence be attributed to a specific intervention, it leaves the recipient free to mix interventions in the spirit of the recent discussion of “combination prevention” [21,22].

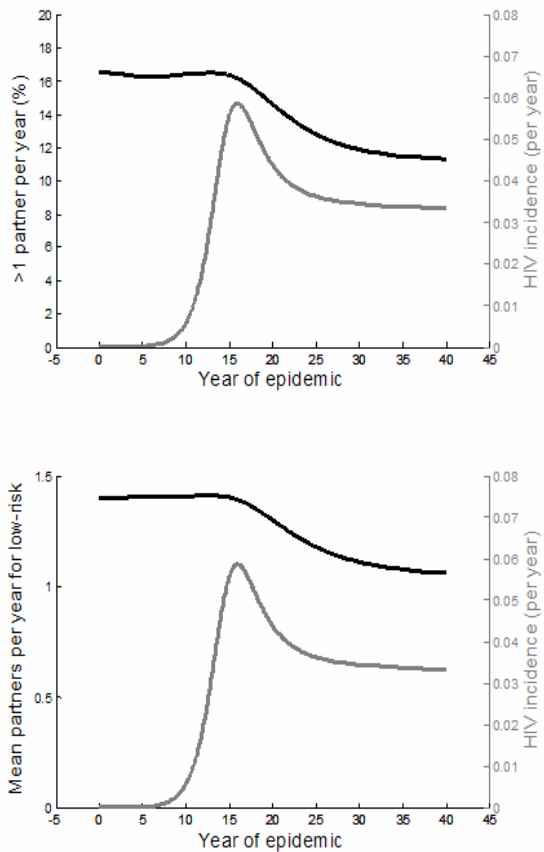
Nor does the COD approach specify how any prize would be distributed among the local actors who have contributed to winning it. We expect that a recipient government might best leverage a COD agreement into HIV prevention achievements by openly discussing the prize distribution with constituent local governments and civil society groups. Depending on the recipient's perception of where incentives would do the most good, the recipient might offer to distribute portions of the prize to local government health officials or to NGO's or even to the mayors of towns. Although the COD contract itself would make no stipulation about how the prize would be spent, a minister of health wishing to use the prize to motivate local governments and their constituents might promise that the award would pay for a football stadium in the towns with the best HIV-prevention performance. The recipient government might even choose to distribute the prize as small payments to individual citizens, an approach that would emulate the incentive structure implicit in the Vanuatu malaria elimination story in the box at the beginning of this essay.

The cash-on-delivery approach is crucially dependent on being able to determine whether HIV transmission has been reduced, and in a fair and timely way. This is harder than determining the impact of other types of programs (such as measuring the progress of antiretroviral treatment programs by counting numbers on treatment) since it amounts to counting non-events (infections averted) and is thus esoteric, invisible and open to debate.

Donors typically assess HIV prevention programs based on measured levels of program activity, on surveys of sexual behaviour or on estimates of the total number of infected people. But the link between program activity and a real effect on transmission is not straight-forward [23]. The metrics of program activity (e.g. number of condoms distributed, testing and counselling sessions attended) do not necessarily indicate any change in individuals' sexual risk behaviour. For instance, most people do not substantially change their behaviour after counselling (and some even increase risk behaviour) [24,25], so the expected impact on the epidemic is minimal [26]. Similarly, reported sexual behaviour may not perfectly correspond to actual sexual behaviour. The reasons for this include the contamination of the information with social-desirability biases (where survey respondents are less likely to report unprotected sex or many casual sex partners [27]) and the difficulties in capturing a complex pattern of behaviour with just a few behaviour indicators[28,29]. For instance, a measurement of 'condom use at last sex' does not record how consistently condoms were used in that relationship, whether or not they were used more at the start of the relationship

or whether condom use would be different with a different partner, all of which determine the net impact of 'condom use' [30].

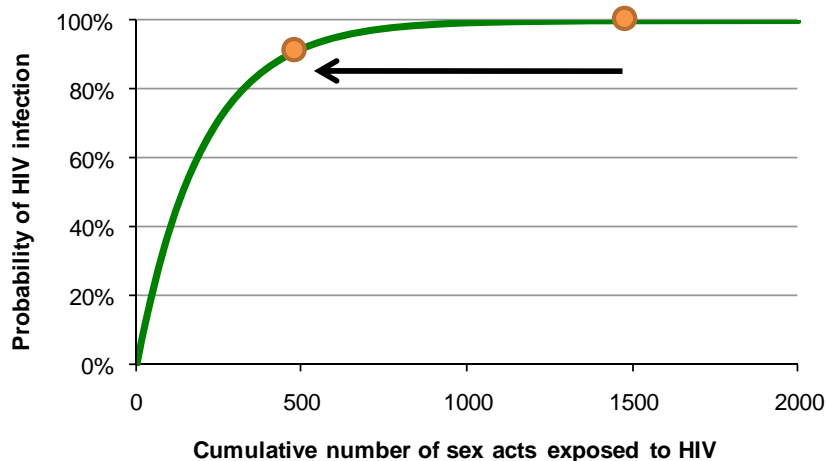
Just as behaviour can change without affecting HIV transmission, HIV transmission can change without a change in individual behaviour. Mathematical models of HIV transmission have shown that, at the aggregate level, indicators of sexual behaviour can change over the course of the epidemic without any individual-level changes [31]. Since AIDS death disproportionately affects those with the riskiest sexual behaviour, over time the population is selectively depleted of individuals with many partners or who do not use condoms, bringing the average measure of these behaviours to 'safer' levels (Figure 1).



**Figure 1:** HIV-related death disproportionately affects those with highest risk so that aggregate-level indicators in sexual behaviour (% with more than one partner in last year or mean partners per year among those in the lowest risk group) can change over the course of the epidemic without change in any individual's behavior. After: Hallett *et al.* 2006 [31].

If there are genuine changes in sexual behaviour, then their impact on transmission is still not clear [23]. Large reductions in risk of one type (e.g. fewer partners) may be offset by another change (e.g. less condom use). It may be possible to calculate an overall measure of risk – for instance, number of unprotected sex acts is influenced by both partner numbers and condom use – but it would not be possible to fully account for all types of behaviour that contribute to the chance of infection. For instance, individuals may reduce partner number and increase condom use, but increase the duration of overlap between partners (increasing concurrency) which can accelerate transmission [32]. Similarly, the choice of partner may be changed, so that the number of unprotected sex acts is unchanged, but the sexual partner is more likely to be infected [28,33]. Hence, it would be almost impossible to fully characterise the true impact that any subtle behavioural shift has on the overall chance of infection acquisition.

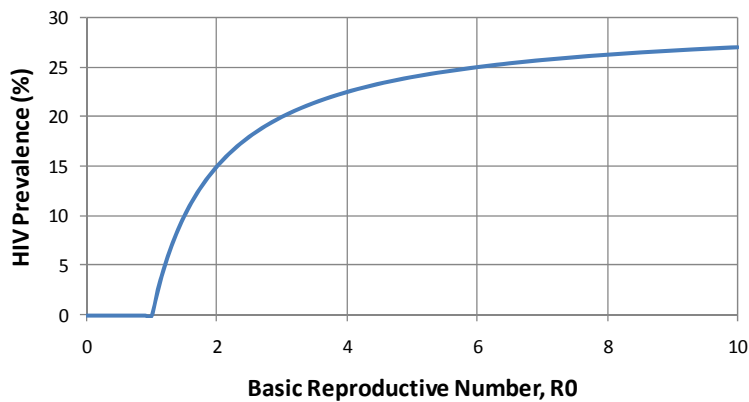
It can also be the case that individuals at high risk of infection can make substantial behaviour change but remain very likely to become infected (Figure 2). On the other hand, behavioural changes among individuals with very low risk – who would be unlikely to be infected anyway – have little or no effect on their own risk of acquisition.



**Figure 2:** Probability that an individual acquires HIV infected as a function of the cumulative number of sex acts exposed to HIV. For highly exposed individuals, substantial reductions in risk (arrow) do not substantially reduce the overall probability of HIV infection. Source: Authors' construction based on the model in the appendix.

Finally, behavioural changes that really do reduce the individuals' risk of infection may still have a limited impact on the whole epidemic. The relationship between the potential for epidemic spread

in a population (summarised in the basic reproductive number,  $R_0$ : the number of new infections generated by one case in a susceptible population ) and the equilibrium level of HIV prevalence exhibits threshold and non-linear behaviour (Figure 3) [34]. If  $R_0$  is high, then substantial behavioural changes among some individuals that makes them fully protected from infection (e.g. perfectly consistent condom use), may not lead to a great reduction in prevalence. Alternatively, if behaviour change is restricted to a part of the population with less risky behaviour where transmission is not sustainable ( $R_0$  less than one), then changes which further reduce the riskiness of their behaviour would have a limited effect on the overall rate of infection.



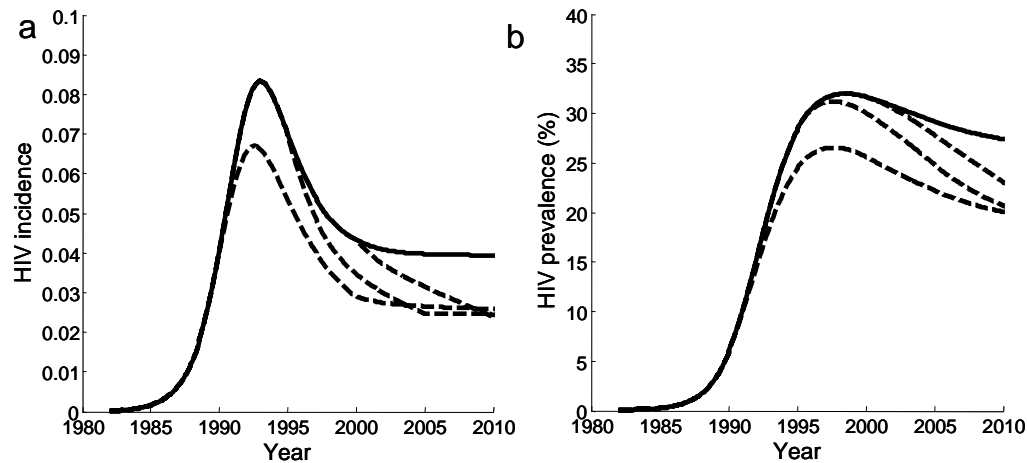
**Figure 3:** The basic reproductive number ( $R_0$ ) determines the endemic level of HIV. (The model population is composed of two sub-populations: one with 70% of the population at low risk of infection and another with 30% of the population at high risk). Source: Authors' construction, using a simplified version of the model presented in the appendix.

For these reasons, it is important to base assessments of the trend in a national epidemic or evaluations of the impact of specific preventive interventions on 'hard' outcomes – reported sexual behaviour indicators can change through bias or the natural course of the epidemic, and may have little bearing on the rate of HIV transmission. The prevalence of certain sexually transmitted infections (STI) has often been suggested a useful "proxy" indicator for the risk of HIV transmission. The relative abundance of STIs and the better correlation between incidence and prevalence for short-duration bacterial infections makes this idea superficially attractive. However, using STI proxies is not reliable since the way in which different sexually-transmitted infections respond to patterns of risk and change in behaviour varies substantially and is determined by their various natural histories. For instance, the prevalence of some STIs is tightly linked only to the behaviours of those at the



highest risk of infection, while the transmission of HIV in mature generalized epidemics is thought to be most linked most to the majority of the population at lower risk. Thus, trends in an STI proxy do not necessarily correspond to changes in risk in the population that will translate to reductions in HIV transmission.

Hard outcomes related directly to HIV are thus the only acceptable measure of impact -- in particular HIV prevalence or HIV incidence. The HIV prevalence rate is the fraction of a given population (typically men and women aged 15-49 years) that is infected with HIV at a given point in time, whereas the HIV incidence rate is the fraction of an uninfected population that becomes infected over a given interval, typically a year. Prevalence is therefore determined by the balance between the incidence of new infections and HIV-related deaths. The heterogeneity in risk of acquiring and transmitting HIV means that there could naturally be a trend in incidence over time [31,35,36,37]. Mathematical models show that incidence is expected to rise quickly at first as the infection spreads among those with the riskiest sexual behaviour but then decline as infections then spread among those at lower risk (Figure 4(a)). Since HIV leads to death after a decade [38], there is also a longer-term reduction in incidence because the individuals with the riskiest sexual behaviour are selectively removed from the population (Figure 1). This can reduce incidence both directly, because there will be fewer high-risk individuals to spread the infection, and an indirect effect, and indirectly, because low-risk individuals will have less opportunity to form risky sexual partnerships [39]. The overall effect is that incidence rises quickly early in the epidemic and falls to a lower level later on. The long interval between HIV infection and AIDS deaths means that the surge in mortality caused by the early spike in incidence will not occur until after incidence has settled at a lower level. This means that for a short time in a mature epidemic, AIDS deaths can exceed current incidence levels and lead to a decline in prevalence (Figure 4(b)). After that wave of mortality, and in the absence of further changes in incidence, prevalence will remain constant. However, in recent years availability of antiretroviral therapy (ART) has increased dramatically [40]. This extends survival for HIV-infected individuals [41], which would naturally lead to an increase in HIV prevalence over time, even if incidence does not change [42].



**Figure 4** Model projections of (a) incidence (b) prevalence for a generalised HIV epidemic without antiretroviral therapy in the absence of behaviour change (solid lines) and with behaviour change (50% reduction in partner change rate) starting in 1990, 1995 and 2000. Source: [5]

It is also possible that increases in AIDS mortality lead not only to reductions in risk in the population (as assumed above) but also to some *increases* in risk. This would occur, for instance, through mortality leading to higher rates of widowhood [43]. Widow and widowers are likely to be infected with HIV (having probably been infected by their late spouse towards the end of their life) but many resume sexual activity with several new partners.

Thus interpreting trends in prevalence is problematic, and it would be much easier and less controversial to use a measure of incidence, if possible. In a mature epidemic, changes in individual behaviour are much more likely to be followed by changes in incidence than by changes in prevalence (Figure 4), and incidence is not confounded by the scale-up of the antiretroviral treatment. Until recently incidence could only be measured in cohort studies where the same individuals are tested for HIV, typically at intervals of 1-2 years. However, these measurements are exposed to an important bias arising from follow-up being more likely among stable populations in individuals, which may be a lower risk of infection than others [44]. Furthermore, the process of repeatedly testing (and having results returned) may itself make the studied population cohort unrepresentative of the general populations [45,46].

It is also important to remember that all measurements of incidence and prevalence are associated with some degree of uncertainty due to sampling error. This means that no estimation of the

difference between end-of-period incidence and what it might have been can be perfectly precise (even if systematic biases, as discussed above, are minimised). In randomly-controlled evaluation trials or routine assessment of national programs, a conservative approach is usually adopted whereby the chance of falsely concluding that behaviour change has reduced transmission is minimised, but in the COD format this could lead to success going unrewarded. In considering approaches that could be used to measure reduced HIV transmission, issues of precision and confidence in the estimation of the number of averted cases are of primary interest.

## **Aim for this Paper**

To explore possible strategies for measuring and rewarding success in preventing new HIV infections for the purpose of implementing a cash-on-delivery program for HIV prevention.

## **The Context and Two Candidate Approaches**

Suppose that a donor and recipient government negotiate and come to agreement on a cash-on-delivery contract which specifies that at the end of a specified time period the donor will pay to the recipient government a reward which is a function of the estimated reduction in HIV incidence and of the precision with which that reduction has been estimated. According to the terms of that contract a baseline population survey is conducted at time  $0$  and a follow-up population survey is conducted at time  $T$ , at the end of the cash-on-delivery contract. The two surveys follow exactly the same methodology, including sample size, mode of sampling, conduct of interviews, supervision and auditing arrangements, HIV testing procedure and so on. The financing of both of these surveys must be arranged at the same time as, or prior to, the signing of the donor-recipient contract.<sup>1</sup>

The two most promising ways to estimate the reduction of HIV incidence are as follows:

1. The “prevalence modelling” approach. Compare the observed trend in prevalence between the baseline and follow-up survey with a counterfactual projection of the prevalence that would have been observed at follow-up incorporating data on the uptake of antiretroviral therapy and assuming no other change after time zero in the historical trends in incidence and mortality.
2. The “tests of recent infection” approach. Determine trend in incidence using Test of Recent Infection (TRI).

Further details on each of these two ideas are provided in the following sections.

### ***1. Compare observed trend in prevalence with a counterfactual projection***

HIV prevalence can be measured in cross-section by both the baseline and the follow-up population surveys. The difference in prevalence between the two surveys is the starting point for determining

---

<sup>1</sup> See Birdsall, Savedov & Mahgoub (2010) for more discussion of the political economy, organizational and institutional issues aspects of the cash-on-delivery foreign assistance strategy and of how it might be applied to the education sector.

whether incidence has declined. But a statistical test for a reduction in prevalence is not sufficient, since, as discussed above, increases or decreases in prevalence could occur even if incidence is unchanged or moves in the opposite direction[37]. To overcome this difficulty, a mathematical model is used to construct a counterfactual projection for HIV prevalence is the *absence* of any behaviour change between the two surveys.

The model used to make the projection should be a mechanistic representation of the spread of HIV [33,47]. If the population of interest is in Africa, then a basic description of heterosexual transmission should be captured. If the population of interest is in, for instance, Eastern Europe, then the mechanism should represent transmission by sharing intravenous drug injecting equipment as well as sexual contact [48]. The model should have parameters with “real-world” meaning (as distinct from a purely phenomenological model) so that data can be used to specify parameter values. The model can be analysed in a Bayesian melding framework to allow the projections and conclusions drawn to fairly reflect the uncertainties in model parameters as well as to synthesise all available information about model inputs and outputs [49]. In this framework, instead of running the model once with one set of fixed parameter values, we run it millions of times, each time with parameter values drawn randomly from ‘*prior*’ distributions that represent knowledge about that aspect of the model. For instance, the parameter for the transmission rate of HIV in each unprotected sex act is from a metaanalysis from recent literature reviews [50,51], with the location and shape of the distribution selected such that 95% of its mass lies within the 95% confidence limits from the metaanalysis. . Where the existing literature is insufficient to provide a prior distribution, an uninformative prior captures this lack of information. For example, in this hypothetical country almost nothing is known about the degree to which individuals with given risk behaviour choose sexual partners with the same or with different risk behaviour. The parameter in the model that captures the degree of mixing is bounded by definition within the closed unit interval [29,52]. For this parameter, the appropriate uninformative prior is a uniform distribution between 0 and 1.

The statistical likelihood of the model runs is quantified based on consistency with available time-series prevalence data [53,54]. In Africa, the richest source of time-series data is from the antenatal clinic (ANC) surveillance system. Clinics are selected for inclusion in the surveillance system and report the HIV prevalence of approximately 300 women attending consecutive appointments within a defined time period [55]. Since different sites will serve different types of populations, across

which HIV prevalence may vary, the data is modelled using fixed effects for each site. Information on the course of HIV prevalence early in the epidemic (before the ANC surveillance system was set up) can be used to further discriminate between model runs. Depending on whether the earlier pre-surveillance information comes from large representative sample surveys or only from convenience samples and expert opinion, the prior distributions for the parameters describing the early history of the epidemic can be empirical or subjective. Information on the use of antiretroviral treatment in the population must be incorporated into the model analysis, since, as discussed, this can lead to HIV prevalence being inflated, and could mask reductions in incidence. In the analysis above, it was assumed that individuals on ART could be identified. Laboratory assays exist for this purpose [56] and the [2008 HSRC South Africa survey](#) tested all individuals for the presence of ART in their blood [57]. If that is not possible, ART coverage can be indirectly estimated through questions on the survey instrument, as was done in the recent [Kenya AIDS Indicator Survey](#) [58]. If self-reported ART is used in the analysis, this should enter in the form of prior information, with a range of uncertainty that reflects the potential for misreporting, rather than as data. Routine surveillance of program activity can be used as a further indirect way to measure ART uptake, which could be used to specify that prior distribution.<sup>2</sup> A full proposed technical methodology is presented in Hallett *et al.* [37]. Details of the simulation model are provided in an appendix to this document.

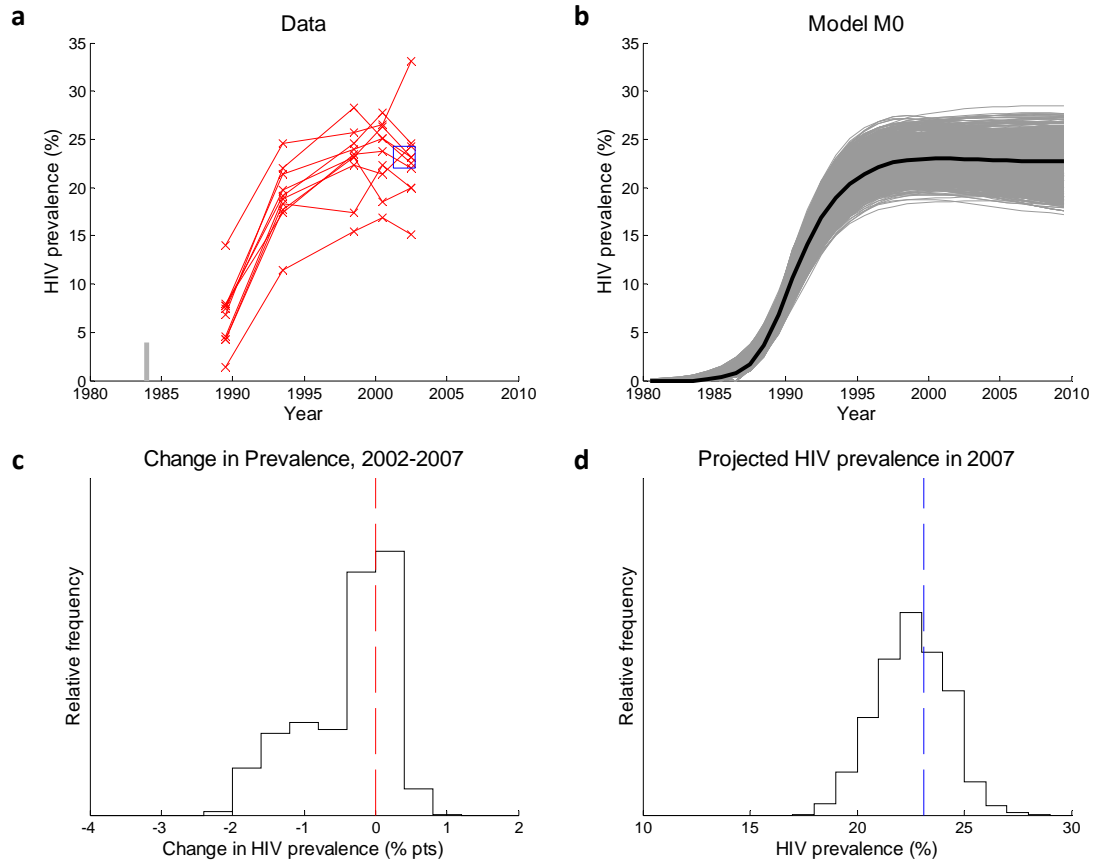
Figure 5 shows an example application of the method using simulated HIV prevalence data for a hypothetical country experiencing a severe HIV epidemic, with prevalence in year 2002 of 23% (Figure 5(a)). In this example, data from the antenatal clinic surveillance system, the prevalence in the baseline survey (with sample size 8000), local data about sexual behaviour and data from the scientific literature about HIV natural history [38,50,51,59,60] have been used to generate a set of model “fits” in the manner described above (Figure 5(b)). It was assumed that there were 10 ANC clinics, each of which sampled 300 women per year. There was a fixed variation between the sites that was normally distributed with mean equal to the true average prevalence plus  $D$  and standard deviation equal to 0.002. This variation is typical of the variation in real ANC datasets, and is due to clinics in the surveillance system being located in different types of populations (e.g. near rural town,

---

<sup>2</sup> A further elaboration of the prevalence modelling method disaggregates each of two sample surveys by age group. Assuming the two surveys are, say, five years apart, then the people aged 20 to 25 in the second survey can be thought of as a random sample from the same cohort as the group that was aged 15 to 20 at the time of the baseline survey. After adjusting for the expected death rate of that age cohort and other possible joiners and leavers, the increase in prevalence between the two surveys is an estimate of the incidence rate in the five-year age cohort. Although a useful tool for diagnosing the age profile of risk in the studied population, for present purposes this approach has the disadvantage of providing only a single estimate of HIV incidence (by age group) and thus does not contribute to the estimation of HIV infections averted.

need main road, etc.) with higher or lower average levels of HIV prevalence. The value  $D$  represents the extent to which the ANC prevalence, across all sites on average, over or under-estimates HIV prevalence in the overall general population. There is evidence that ANC estimates tend to be somewhat too high [61], which may be due to pregnancy, especially at younger ages, being correlated with higher risk of HIV [62]. In that case,  $D$  can be set to a positive value, estimated from a comparison of ANC prevalence estimates and population survey estimates. For simplicity in this demonstration,  $D$  was set to zero.

The variation between the models runs reflects the inability of the existing data to uniquely specify just one set of parameters as being consistent with the data. That is, uncertainty in many factors that contribute to the spread of HIV means that several sets of parameters are supported by data. This uncertainty leads to a range of projections for HIV prevalence for the period 2002 to 2007. Even without changes in sexual risk behaviour, most of the simulations project a decline in prevalence between 2002 and 2007— typically by 1-2% over the five year interval (Figure 5(c)). The range of prevalence values that could be expected in 2007 is thus, 17% to 28% (Figure 5(d)) (the wider limits also reflecting between-run variation due to sampling errors in the prevalence measurements). Essentially, it is this distribution of projections for HIV prevalence in 2007 under the “null assumption” of no behaviour change that is used as the counterfactual to the “observed” prevalence in the follow-up HIV prevalence survey. Thus, it can be seen that a measured value of 20-21% HIV prevalence in 2007 – although actually a substantial decline on the 2002 measurement – would still be entirely consistent with this null projection, and there would not be strong evidence for behaviour change having affected the course of the epidemic. However, HIV prevalence measurements at follow-up in 2007 smaller than 17-18% would be very unlikely if there had not been changes in behaviour reducing the rate of HIV transmission, and thus measured 2007 prevalence this small would be evidence that behaviour change had affected the course of the epidemic.



**Figure 5: Projection of 2007 HIV prevalence based on data through 2002 under the assumption of no subsequent behaviour change (Model  $M_0$ ).** (a) The data from ten antenatal clinics (red cross and lines: each cross shows one point estimate from one clinic and the lines joins estimates from the same clinic) and the baseline survey in 2005 (blue square). The prior limit on prevalence in 1984 is shown as a vertical grey bar. (b) Model projections assuming no behaviour change (called Model  $M_0$ ): each thin grey path shows one item in the posterior distribution and the thick black path shows the best supported projection. (c) Distribution of the expected within-run change in prevalence between 2002 and 2007 using Model  $M_0$  (i.e. without behaviour change). The vertical red line shows the naive null assumption of zero change. (d) Distribution of the expected prevalence in 2007 using Model  $M_0$  (i.e. without behaviour change). The vertical blue line shows the prevalence estimate in the baseline survey in 2002.

A statistical test of the hypothesis that risk behaviour changed after 2002 can be formulated as a model comparison between the constrained model without behaviour change (model  $M_0$ ) and the unconstrained model with possible behaviour change (model  $M_1$ ). In model  $M_1$  the possible timing of any behaviour change between the baseline and follow-up survey and the magnitude of the change in risk behaviour are characterized by an uninformative and an informative prior, respectively, the latter based on trends in behaviour indicators measured at the time of the surveys. The Bayes Factor ( $K$ ) is comparable to the likelihood ratio of non-Bayesian statistics and measures the degree to which the numerator model is superior to the denominator model in explaining the newly observed data, across the range of parameter space supported by the prior information (Equation 1). Using Model



$M_0$  as the reference model in the denominator, a sufficiently large  $K$  statistic allows us to judge whether behaviour change has occurred subject to the uncertainty incorporated through the Bayesian priors on parameters and earlier data.

$$K = \frac{\int p(\boldsymbol{\theta})p(\mathbf{W}|\boldsymbol{\theta}, M_1)d\boldsymbol{\theta}}{\int p(\boldsymbol{\theta})p(\mathbf{W}|\boldsymbol{\theta}, M_0)d\boldsymbol{\theta}}$$

**Equation 1:** Measuring the relative performance of the two models  $M_0$  and  $M_1$ .  $p(\mathbf{W}|\boldsymbol{\rho}, M_x)$  is the likelihood of the simulated epidemic given the set of parameters  $\boldsymbol{\theta}$  and model  $M_x$ , and  $p(\boldsymbol{\theta})$  is the likelihood of that parameter set.

It has been proposed [63,64] that the natural way to interpret  $K$  values is on a logarithmic scale, such that:

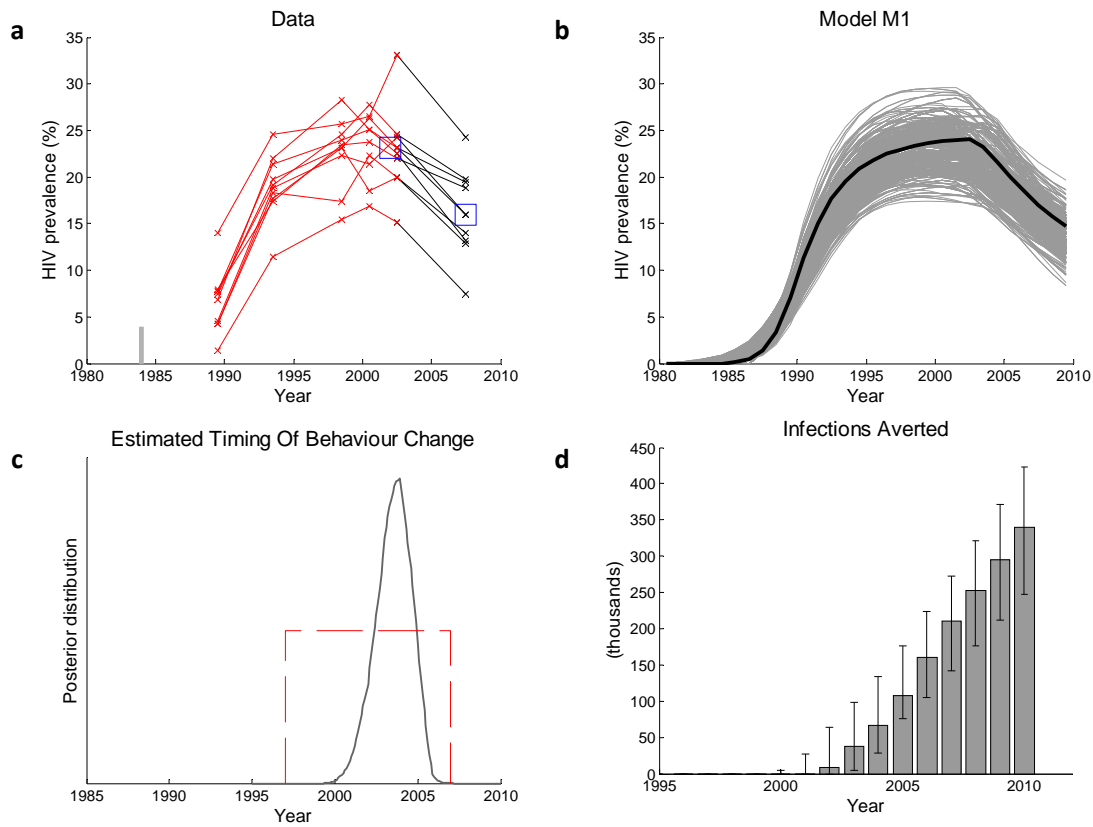
- $0 \leq 2\ln(K) < 2$  indicates little or no support for model  $M_1$  / behaviour change affecting epidemic;
- $2 \leq 2\ln(K) < 5$  indicates some support for model  $M_1$  / behaviour change affecting epidemic
- $5 \leq 2\ln(K) < 10$  indicates strong support for model  $M_1$  / behaviour change affecting epidemic;
- $2\ln(K) \geq 10$  indicates very strong support for model  $M_1$  / behaviour change affecting epidemic.

The  $K$  value would be fed-forward to the payout function to determine the cash-on-delivery reward for the incidence reduction.

Figure 6 demonstrates the application of the prevalence modelling approach to the hypothetical country discussed above, after a baseline survey in 2002 and a follow-up survey in 2007 (Figure 6(a)). In this hypothetical country, prevalence declined between 2002 and 2007 from 23% to 16% as a result of changes in risk behaviour beginning in 2002 (the data were simulated, so in this hypothetical example the “truth” is known).

The solid dark path in Figure 6(b) shows the time trend of population HIV prevalence that achieves the largest value of the  $K$  statistic and thus is deemed to best fit the updated simulated data. Note that the 16% value given by that best fitting path for HIV prevalence in 2007 is outside the range of

values supported under the null model (Figure 5(b) and 5(d)). Thus, the fit of Model  $M_1$  incorporating changes in risk behaviour (Figure 6(b)) is much better than that of the constrained model that does not allow behaviour change. In this demonstration, the computed value of  $2 \ln(K)$  is 72, much larger than the conventional threshold of ten, which would be strong evidence for behaviour change driving this decline in prevalence, rather than natural epidemiological dynamics alone. This finding would fulfil the payment conditions of a COD contract which had specified that payment be made if HIV infections were averted by post-2002 behaviour change.



**Figure 6: Projection of 2007 HIV prevalence based on data through 2002 and allowing for subsequent behaviour change (Model  $M_1$ ).** (a) The data from ten antenatal clinics (red cross and lines: each cross shows one point estimate from one clinic and the lines join estimates from the same clinic) and the baseline survey in 2002 and follow-up survey in 2007 (blue squares). The prior limit on prevalence is shown as a vertical grey bar. (b) Model  $M_1$  (which allows behaviour change): each thin grey path shows one item in the posterior distribution and the thick black path shows the best supported projection. (c) The uninformative prior (dashed line) and updated posterior distribution for the timing of behaviour change (solid line). (d) Estimated cumulative number infections averted (error-bars shows 95% interval) in the population.

Figure 6(c) shows the uninformative prior distribution (dashed red lines) and the posterior distribution (solid lines) for the dates during which behaviour change occurred. (Recall that the prior

distribution describes the analysts' subjective view of possible dates when behaviour change might have occurred before the analysis is run, and that the posterior distribution is a version of the same probability distribution which has been subjected to Bayesian updating in light of the data). The data were originally simulated using a reduction in risk between 2002 and 2004 starting immediately after the baseline survey. The model analysis shows strong support for changes in risk behaviour in that period. The prior (1997-2007) was set to include some years where behaviour did not change (i.e 1997-2002, before the baseline survey), and the posterior shows little or no support for behaviour change in that period. That is, the estimated timing of risk reduction coincides with when the behaviour change was in fact introduced into the model, and this temporal resolution contributes to the evidence for the detected changes in risk being associated with the launch of the COD agreement.

It is also possible to calculate the number of infections averted in the model  $M_1$ , with the posterior distribution of model fits providing the basis for 95% credible intervals (Figure 6(d)). This information on the most likely number of fewer infections in 2007 and on the precision with which that number is estimated serves as input to the payout function which determines rewards.

It should be noted that, strictly, the data used to monitor trends before and after the baseline could be the antenatal clinic data or only the baseline and follow-up surveys, as well as a combination of the two, as shown in the example. (However, the ANC data is always required to inform the historic trajectory of the epidemic before the baseline survey, aggregated measures of which are already publically available [40]). The advantage of relying more heavily on the baseline and follow-up surveys are that: (i) Household-surveys are thought to provide a closer approximation to general population samples, and thus a less biased estimate of the overall scale of the epidemic [61,65]; and, (ii) observed changes in prevalence are not partially confounded by changes in patterns of fertility, or clinic attendance, which could potentially generate spurious trends in ANC prevalence estimates [47,55,66].

## **2. Determine trend in incidence using Test of Recent Infection (TRI).**

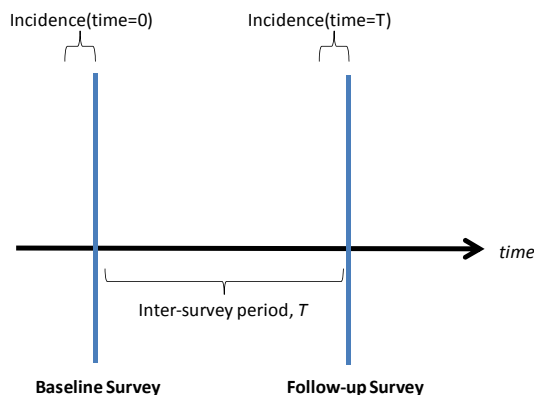
A direct measure of incidence would provide a better tool to plan and evaluate HIV programs [67]. The most direct method is to follow a large cohort of HIV negative people over time, as is done, for example, in Phase III HIV vaccine trials, but cohort studies are prohibitively expensive and often unrepresentative [46]. Although the prevalence modelling method described in the preceding section can be used to estimate incidence and averted HIV infections [68,69], at least three surveys would be required to directly estimate changes in incidence and modest breaches in some in-built stability assumptions could generate misleading results [68]. Therefore, a practical and valid method of measuring incidence from cross-sectional surveys would be ideal, and has been a goal of testing research for over a decade. Recently a number of new testing algorithms have been developed in the hope of serving this purpose [70].

The underlying principle of a set of new blood tests, or Tests For Recent Infection (TRI), is that the immunological response to HIV evolves over the first months of infection, and by measuring the quantity, proportion or avidity of HIV antibody, recent infections can be discriminated from older ones [71] (Figure 8). The most widely used of these assays is the BED capture enzyme immunosorbent assay ('BED test'), in which the optical density varies according to proportion of IgG that is anti-HIV antibody [72]. A test is "BED-recent" if the test produces a spot that is sufficiently translucent to have a measured normalised optical density of less than 0.8; if it is more opaque, it is "BED-chronic". Earlier analytic approaches had considered that a BED-recent results means that infection had occurred sometime in the previous 150-187 days [73,74,75]. Thus, the rate of new infections could be estimated because the question of incidence estimation becomes only a question of counting the number of individuals in that category [73].

BED-derived estimates of incidence have been compared with gold-standard measures of incidence in a range of settings [73,76]. The common finding has been that BED estimates are substantially too high [76,77], leading to calls for caution in the use and interpretation of the test [78]. It has become clear that this is because the test misclassifies some individuals infected for a long time as being "BED-recent" [77]. If the proportion of such misclassifications were constant across settings or known for an individual setting, the incidence estimators could simply be deflated by the appropriate proportion to correct for these misclassifications [79] [80] [76,77,81,82,83]. Unfortunately, the fraction of individuals infected for more than a year that are misclassified as

recent varies between 1.7% (South Africa[82]) and 27% (Rwanda and Zambia[76]), casting doubt on the reliability of any HIV incidence estimates which rely on such a simple adjustment [84]. There are indications that this variability results from individuals with chronic infection regressing back to the BED-recent state as the immune system deteriorates [76,83,85,86,87], and from the differential survival of those who do not develop a sufficient immune response to ever be BED-chronic (who tend to be elite suppressors and survive for longer) [56,88] compared to others. This means that the values needed to correct for misclassification vary by age-group, over time and between countries [84], leading to difficulties in drawing comparisons and measuring trends. Thus, neither absolute levels or trends in incidence can be reliably determined by the BED test alone [89].

Nevertheless, because of recent theoretical advances, the TRI approach can now be proposed as an alternative, potentially superior, approach to measuring averted HIV infections. The new modelling work shows that applying several blood tests to each individual blood specimen could reduce the percent of mistakenly identified recent infections to less than 2 , provided the sample size of the survey data were sufficiently large [89,90]. Knowing that misclassifications are most likely for individuals whose immune systems are either extremely effective or have been depleted by long infection, blood samples identified by the BED as recently infected can be re-tested for avidity (to screen out those with exceptionally resistant immune systems) and for ART or viral load (to screen out those with depleted immune systems). Such a combined testing strategy has been shown to reduce the rate of recent-classification among those infected for more than a year from more than 16% to 1% [71,83]. As applied research to test such a combined strategy in the field goes forward, we can prepare for its success by considering the TRI approach as a promising measurement strategy to implement the COD approach to HIV prevention [f,g].



**Figure 9:** Design for measuring a change in incidence using the tests for recent infection technology.

To measure the change in incidence over a period, the TRI approach estimates incidence in the period immediately prior to the baseline and follow-up survey and subtracts the second from the first (Figure 9). Incidence in each period is estimated as:

$$I(t) = \frac{U(t) - \varphi O(t)}{H(t)E[\tau_R]}$$

**Equation 2:** Estimator of incidence,  $I(t)$ , using test for recent infection technology [79].  $U(t)$ : HIV-infected & Under-threshold;  $O(t)$ : HIV-infected & Over-threshold;  $H(t)$ : Not HIV-infected;  $E[\tau_R]$  is Expected time in recent state (for those that do progress); and,  $\varphi$  is the odds of those of being in the non-progressing sub-population.

When comparing incidence rates at the baseline and the follow-up surveys, it is not necessary to know the value of  $E[\tau_R]$ , and if the odds of non-progression are zero, the statistic becomes simply a comparison of the ratio of individuals classified as recent to HIV-uninfected individuals in each of the surveys (Equation 3).

$$IRR = \frac{I(T)}{I(0)} = \frac{U(T)/H(T)}{U(0)/H(0)}$$

**Equation 3:** Under certain conditions, the comparison of incidence between two time-points collapses to a comparison of two proportions. IRR: Incidence Rate Ratio.

### **Confidence, Time between Surveys and Sample Size Considerations**

The preceding sections have discussed provisions that can be taken to safeguard the prevalence modelling approach and the TRI approach from the systematic biases to which they are inherently vulnerable. Since both approaches rely on HIV tests from two cross-section surveys conducted at the baseline and a few years later, it is also important to consider the role of random statistical sampling error. These errors can reduce the chance that there is sufficient evidence to justify reward payment, or reduce the amount of the reward, even when HIV infections have truly been averted. These statistical errors can be minimised by increasing the sample size of the baseline and follow-up

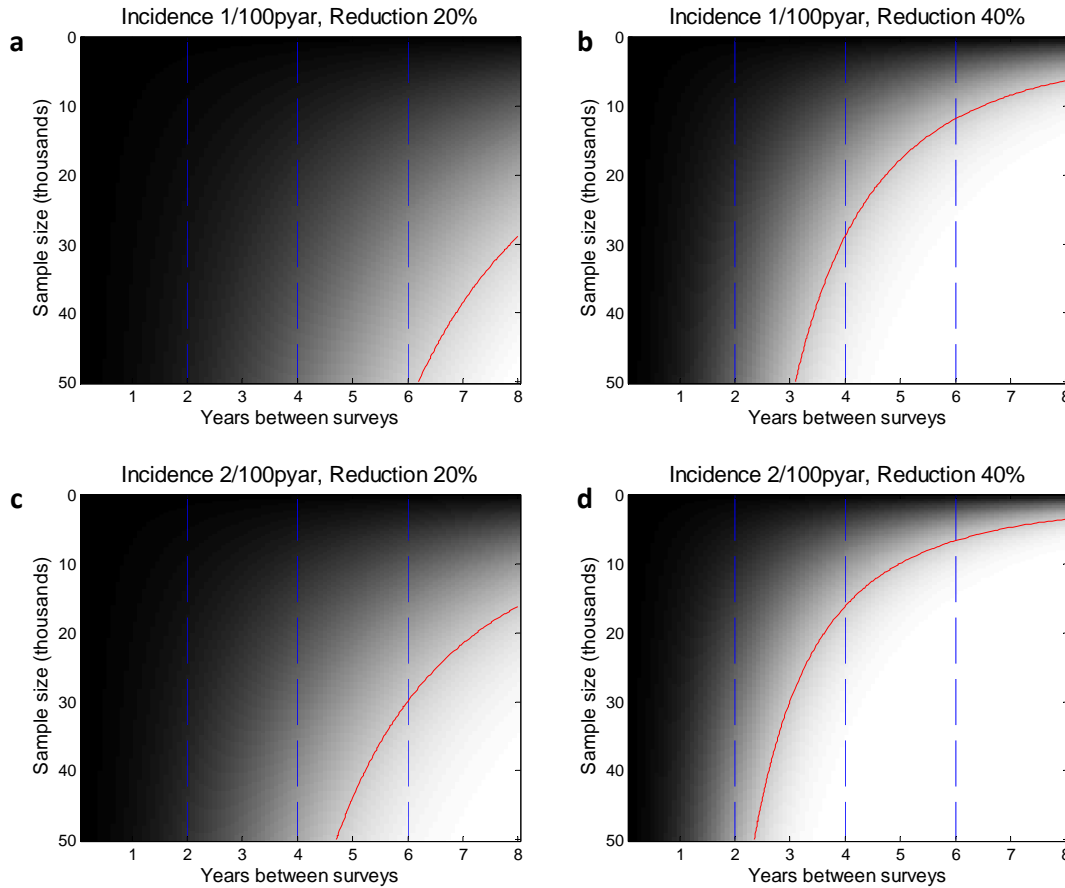
surveys, or by increasing the interval between the surveys. But both of these remedies are costly: larger sample sizes are more expensive to implement and longer intervals between the surveys mean that countries have longer to wait before the results are available, the payment is made and planning for the next round of interventions can start.

The trade-off is illustrated in Figure 10, where an idealised TRI is used to estimate incidence at the baseline and follow-up surveys. In principle the same trade-off applies if the prevalence modelling approach (discussed above) is used instead. In the TRI design, a simple criterion for reward payment might be the measurement of a statistically significant reduction in the estimated incidence rate (at the 5% level). In the figure, panels (a) and (c) depict the tradeoff between sample size and survey interval required to detect an improvement when the true five-year reduction in incidence has been 20 percent or four percent each year, while panels (b) and (d) do the same for a true five-year reduction of 40 percent or eight percent each year.<sup>3</sup> The red line in each panel shows the trade-off between sample size and survey interval to assure an 80 percent chance of recording a significant difference, with a 20 percent chance that an incidence reduction would be missed (i.e. to assure the test has 80 percent power). Moving to the southeast in any of the four panels, which represents increasing either the sample size or the survey interval, increases the probability of recording a significant difference when one has occurred (i.e. increases the statistical power of the test) as represented by the lightness of the shading. For instance, if incidence is initially 2/100 person-years at risk (pyar) and the intervention truly does reduce incidence at a five-year rate of 40 percent (panel d), then a reward criterion with 80 percent power can be achieved with a survey interval of 2 years and a sample size of ~40,000 individuals. Increasing the interval to 5 years means that only ~10,000 individuals per survey would be required. In populations with lower incidence, larger sample sizes are required: for instance, if incidence is 1/100pyar and incidence is reduced at a five-year rate of 40 percent (panel b) over a five-year survey interval, then sample sizes of 15,000 would be required to achieve 80 percent power. Greater sample sizes and/or longer intervals between surveys are required if the expected reduction in incidence is more modest: for instance, in a population with 2/100pyar incidence and incidence decreases at a five-year rate of 20 percent (panel b), then a sample size of 40,000 and inter-survey interval of 5 years would not be sufficient to provide 80 percent power. For populations with low incidence where incidence reductions are expected to be

---

<sup>3</sup> If the true incidence reduction between two surveys five years apart is 40 percent, we assume that the true incidence reduction between surveys T years apart would be  $T * 40/5$  percent. This linear interpolation/-extrapolation of our assumed rate of incidence decline is a compromise between assuming a constant percentage decline, for which the absolute decline would be largest the first year, and a cascading decline, for which the absolute decline would be smallest the first year. See Hallett [2009] for further discussion.

modest (panel a), it is difficult to create a design with sufficient power – for instance, surveys of 50,000 individuals separated by 6 years would be required to generate 80 percent power if incidence is 1/100pyar and incidence is reduced at a five-year rate of 20 percent. This practical problem can be addressed by considering alternative reward criteria or “payout functions”.



**Figure 10:** Trade-off between time between the survey interval, and the survey sample size to achieve an equal chance of detecting the reduction in HIV incidence. Panels show statistical power (white indicates close to 100 percent power, and darker shades lower power; the red line is at the power=80 percent iso-cline). The blue dashed line indicates intervals between surveys of 2, 4 and 6 years. The reduction in incidence is simulated as a steady decline in incidence, with reductions over 5 years equal to (a,c) 20% or (b,d) 40%.

### Criteria for selecting a mutually agreeable payout function

A “payout function” is defined as a rule that determines whether the recipient party in a COD agreement receives a reward and, if so, what proportion of the total maximum reward is paid. Since at the conclusion of the COD measurement period the donor does not know the true achievement of the recipient, but only the estimated magnitude of that achievement based on the two population surveys, the payout can be based on both the estimated magnitude and on the precision of the



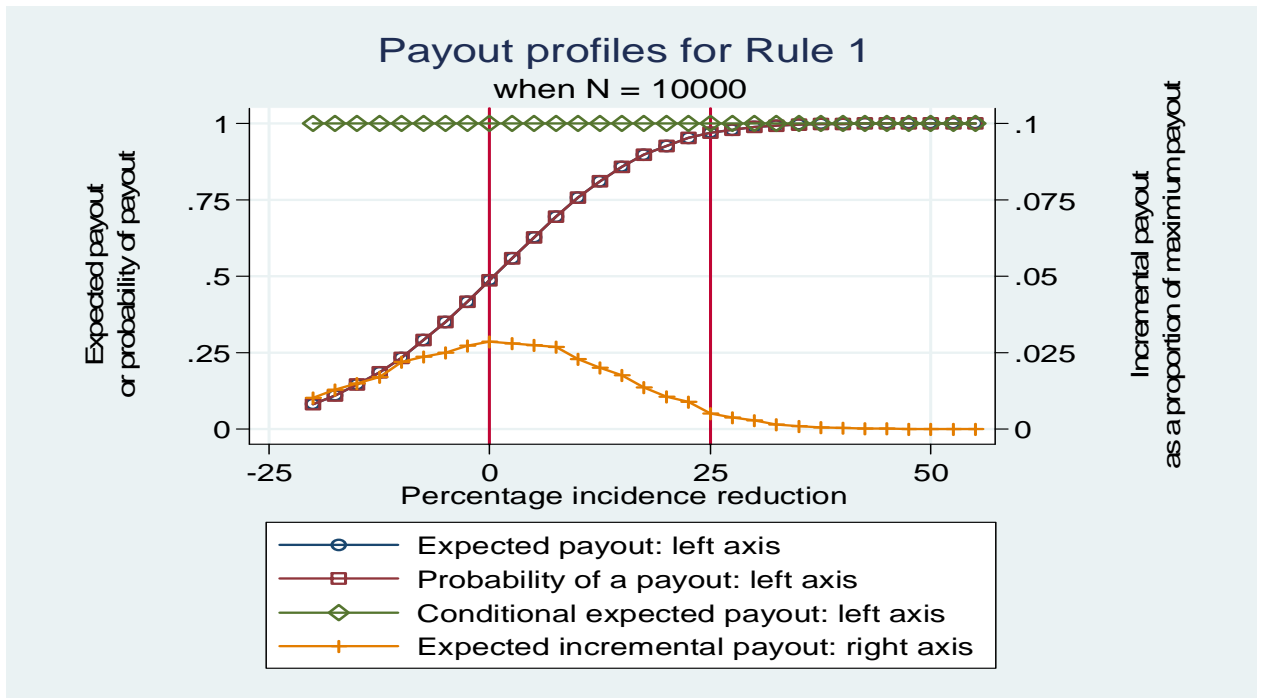
estimate. Rewarding the estimated size of the improvement provides the desired incentive for the recipient government and its constituencies to cooperate effectively in reducing HIV incidence, while rewarding precision assures the donor that the probability and/or the magnitude of the payout will be on average associated with true improvement, not just with a sampling or non-sampling error of measurement. One can imagine making the payout a function of achievement in various subsets of the population, such as improvement among women (an equity criterion) or among high risk groups (an efficiency criterion), but we set aside these possible refinements for the practical reason that they would require larger samples or more frequent survey intervals and for the philosophical reason that improving the broadest possible population average of HIV incidence is consistent with both equity and efficiency objectives and accords with the COD philosophy of leaving the details to the recipient.

An unlimited number of possible payout functions can be constructed which reward the size and the precision of the measured improvement. These differ from one another in a variety of dimensions among which we focus on four that we think will be important to both the donor and the recipient of the COD agreement. These are the profiles with respect to the true incidence reduction of (1) the expected payout; (2) the expected incentive; (3) the probability that a payout occurs; and (4) the conditional expected payout.

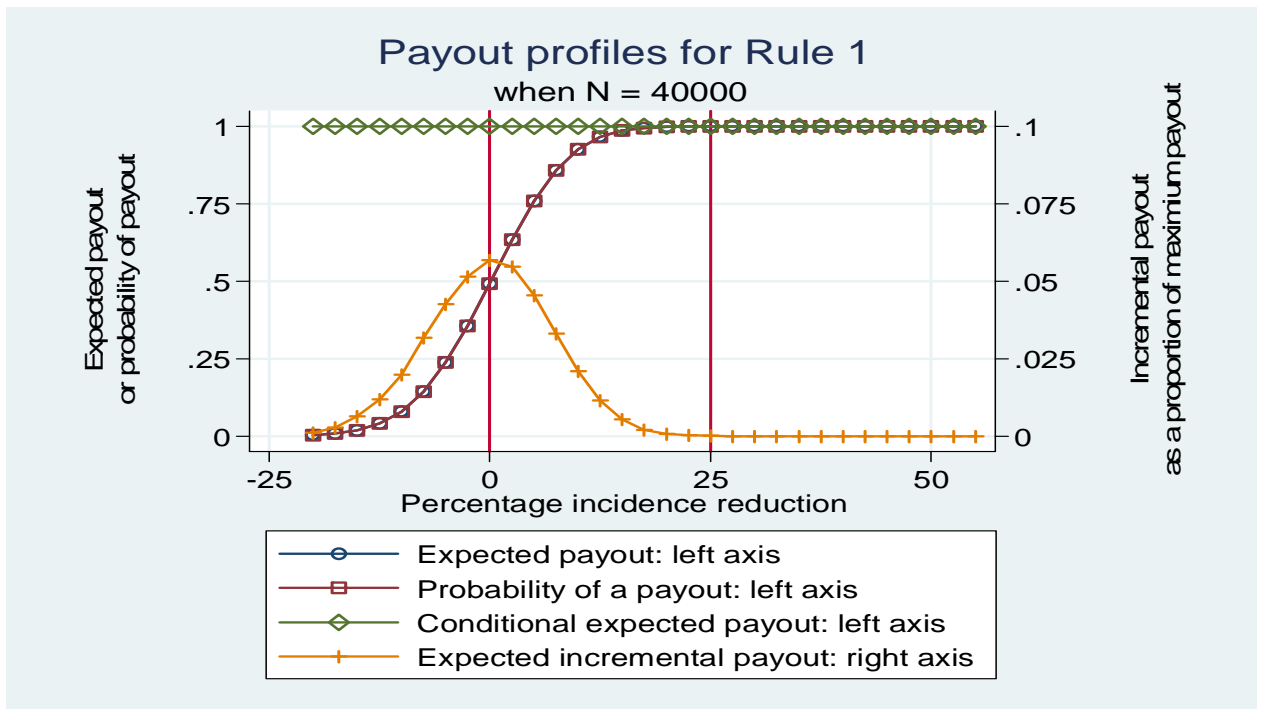
We define the “expected payout profile” of a payout function as the relationship of the average amount paid to the true reduction achieved in HIV incidence. To construct the expected payout profile for a candidate payout function, we use the computer to simulate the accuracy of two population surveys at measuring a true incidence reduction. We perform the simulations over a range of values for the actual incidence reduction, from a failed HIV prevention which increases HIV incidence by 20 percent to a highly successful one that reduces incidence by 50 percent. In all cases we assumed that the incidence at the time of the follow-up survey was 2/100 person-years at risk, or two percent per year, and that the baseline and follow-up surveys each has 20,000 individuals.

For example, suppose the donor and recipient agree that the donor will pay the recipient the full amount of the reward if the estimated HIV incidence from the second population survey is less than that from the first, regardless of whether the difference is statistically significant. Figure 11 displays the four profiles of this simple payout function for two sample sizes: 10,000 households in panel (a) and 40,000 households in panel (b). In both panels, the horizontal axis represents the true reduction in incidence between the baseline and follow-up surveys, with improvements up to a 50 percent reduction towards the right. Values to the left of zero represent a worsening of HIV incidence up to

Panel (a). Payout profiles for Rule 1 when survey sample size is 10,000



Panel (b) Payout profiles for Rule 1 when survey sample size is 40,000



**Figure 11.** Four simulated profiles of the payout function which awards the recipient if any measured reduction in incidence has occurred, regardless of its magnitude or statistical significance. Panel (a) Sample size is 10,000; Panel (b): Sample size is 40,000.

a 25 percent exacerbation. The first three profiles use the left vertical axis, while the expected incremental payout uses the right. Each point on each line is generated by 100,000 random draws of the appropriate number of households under the assumption that the true reduction in incidence has the value on the horizontal axis.

Because this simple payout rule resembles a weak version of a classical hypothesis test, it is not surprising that the profile of expected payout for this function displays the classic shape of a statistical power function, an S-shaped curve or ogive. Because the payout is either all or nothing (zero or one), the profile of the expected payout function is identical to that of the probability of payout. (In Figure 11, one of these is superimposed over the other, leaving only three curves in each panel.) These expected payout functions all display one desirable property: they are monotonically increasing from left to right, which means that on average the donor will pay out a larger proportion of the maximum award when the recipient has achieved greater incidence reductions. This feature of the payout functions is captured by the “expected incentive profile”, or the incremental expected payout, which in this case is always positive at all values of the incentive reduction, an essential feature of a COD payout function upon which a donor should insist<sup>4</sup>.

This payout function displays several other features that may make it desirable to the recipient, but unacceptable to donors. First, for both depicted sample sizes, and indeed for all sample sizes, the expected payout profile of this payout function intersects the vertical axis at 0.5. Thus, even though the recipient has achieved no improvement, or even allowed HIV incidence to increase, the recipient’s expected reward (as well as the probability of receiving a reward) can be up to 50 percent. Donors may also be uncomfortable with the fact that when the surveys have erred, the recipient receives the totality of the reward, as represented by the horizontal conditional line at the top of the graph giving the profile of the conditional expected payout. At a higher sample size, the chance is reduced that two population surveys would mistakenly measure a decline in incidence when incidence has actually increased. This is why the S-shaped curves are steeper in panel (b). However, even at the higher sample size, with an actual increase of 2.5 percent in incidence, this function would mistakenly reward the recipient the full amount of the prize about 20 percent of the time. Finally, the hill-shaped incentive profiles in the two panels, though always positive, are centered on the origin and drop quickly to the right, indicating that the recipient is most highly

---

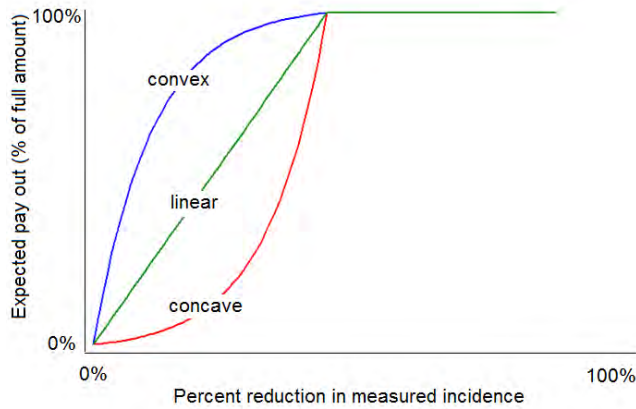
<sup>4</sup> The incentive profile is defined as the slope of, and computed as the partial derivative of, the expected payout function with respect to the true incidence reduction.

rewarded when achieving the initial small reduction in incidence. With the larger sample size, the incentive profile is so tightly concentrated on the origin that the COD would provide little additional reward for going beyond a 5 percent improvement and virtually none for going beyond a 10 percent reduction. By focusing all of the COD incentives on such a small reduction in incidence, payout function 1 may be unattractive to both parties, who would like to see bigger rewards for big improvements.

Table 1 lists ten example payout functions starting with the one already described. The table groups the ten functions into two categories – (i) “Threshold” functions, that pay a fixed amount if the ratio of the new to the old estimated incidence rate is below a certain level or the change is statistically significant; and, (ii) “Continuous” functions, that pay in proportion to the estimated reduction in incidence or, equivalently, in proportion to the estimated number of HIV infections averted up to a maximum reward. For example, the threshold functions might offer a million dollar prize if the threshold is reached, while the continuous functions might offer \$100 per HIV infection averted up to a maximum of a million dollars.

Payout Function		
<b>Threshold Functions</b>	1.	Pay if any reduction in measured incidence rate (irrespective of p-value).
	2.	Pay if reduction is large (>20%) (irrespective of p-value).
	3.	Pay if reduction reaches statistical significance at $p \leq 0.05$
	4.	Pay if reduction reaches statistical significance at $p \leq 0.20$
	5.	Pay 100% if $p < 0.05$ ; Pay 50% if p-value 0.05-0.20
	6.	Pay 100% if $p < 0.05$ ; Pay 75% if p-value 0.05-0.10; Pay 50% if p-value 0.10-0.20; Pay 25% if p-value 0.20-0.30
<b>Continuous Functions</b>	7.	Pay linearly in proportion to the reduction in incidence (up to a reduction by half). (See Figure 12)
	8.	Pay linearly in proportion to the reduction in incidence (up to a reduction by half), with a bonus if the reduction reaches statistical significance at $p = 0.05$ .
	9.	Pay convexly (faster return at smaller reductions) in proportion to the reduction in incidence (up to a reduction by half), with a bonus if the reduction reaches statistical significance at $p = 0.05$ . (See Figure 12)
	10.	Pay concavely (faster increase in return at larger reductions) in proportion to the reduction in incidence (up to a reduction by half), with a bonus if the reduction reaches statistical significance at $p = 0.05$ . (See Figure 12)

**Table 1:** Possible “Payout functions” that could be selected ex ante by negotiated accord between donor and recipient and embodied in the COD agreement which would subsequently govern reward payment as a function of the estimated incidence reduction or the number of averted HIV infections. Full or 100 percent payment is assumed to be the same for all functions.



**Figure 12:** The shapes of alternative continuous payout functions described in Table 1 (functions 7 – 10). The expression used to generate the curves is:  $f = \max(0, \min(1, (\exp(h * ((1 - RR_m) / S_c)) - 1) / ((\exp(h)) - 1)))$ , where  $RR_m$  is the measured incidence rate ratio,  $S_c$  is the value of  $RR_m$  at which there is maximum payout (here corresponding to 50% reduction in incidence), and  $h$  is the shape parameter for the curve (less than zero, the curve is convex; more than zero the curve is concave; approaching zero, the curve is a straight line). Values used for  $h$  are -4 (convex), 0 (linear), and 4 (concave).

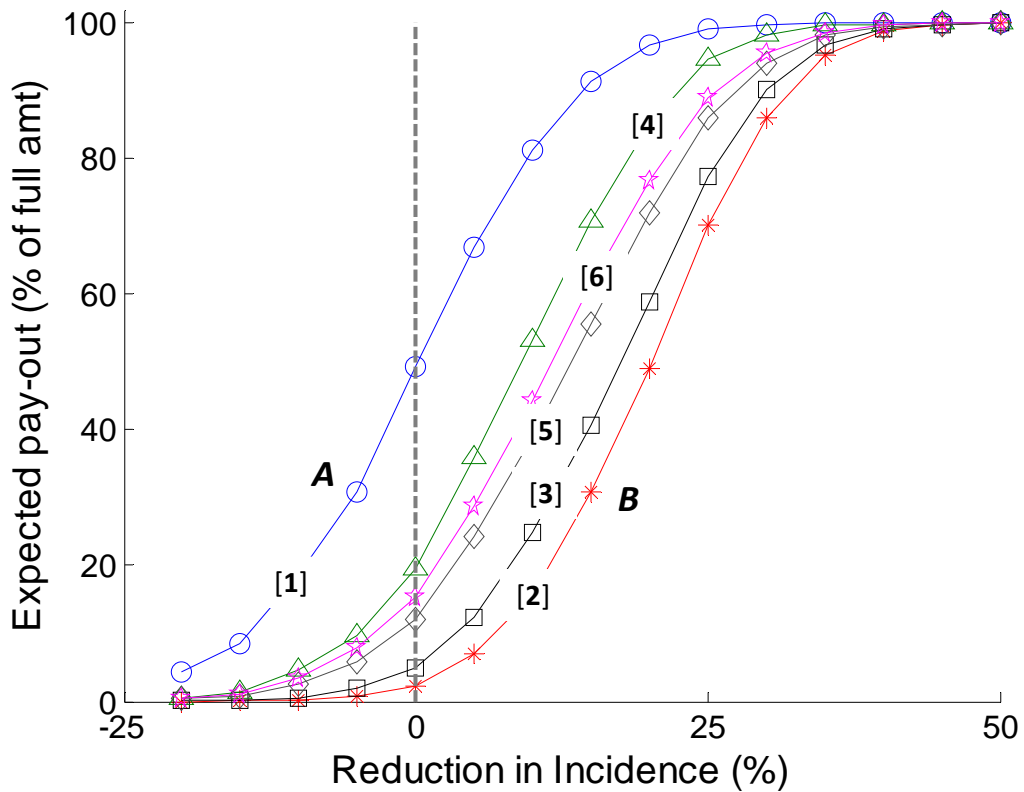
Figure 13 shows the simulated expected payout profile for the threshold payout functions 1 through 6, while Figure 14 shows them for the continuous functions, 7 through 10. As in Figure 11, the horizontal axis shows the true reduction in incidence, and the vertical axis shows the mean or expected payout that would be made on the basis of one of the ten decision rules. The arrows indicate the regions where the mean payout is too low, because the estimated measure of the incidence reduction is less than it really was (“Overpayment”), or the mean payout is too high because the measurement erred in the other direction and exaggerated the country’s achievements (“Underpayment”). Figures 13 and 14 are constructed for a sample size of 20,000, in between the 10,000 and 40,000 sizes shown in Figure 11.

Payout function 2, which is shifted farther to the right than any of the threshold functions, is a more conservative strategy, rewarding a country only if incidence is reduced by 20 percent or more. Its shape resembles that of payout function 1, displayed in Figure 11, but is shifted to the right and centered on 0.2 instead of on zero. Using this payout rule, a country which does not succeed in reducing the incidence rate is rarely rewarded even at the smaller sample size of 20,000. But, another consequence of using the scheme is that some countries that successfully reduce incidence are under-rewarded (e.g. countries that reduce incidence by up to 20 percent have no more than a fifty percent chance of receiving the full payout) and this problem is worse at higher sample sizes.

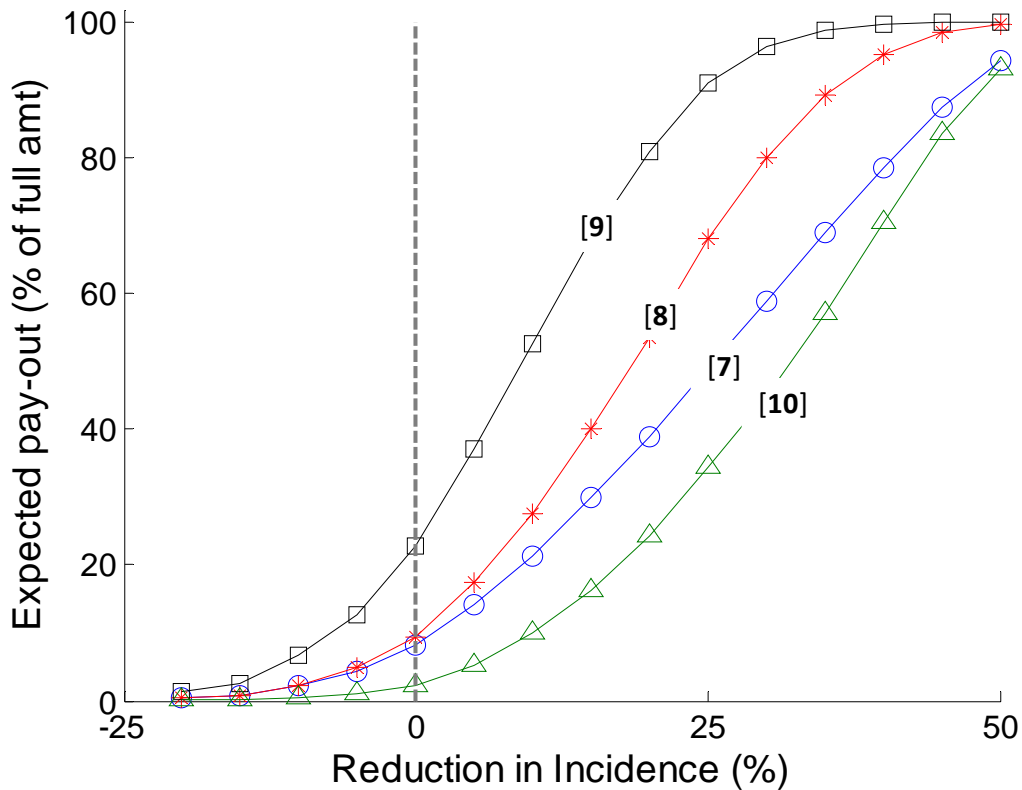
Payout function 3, which is based on a classical hypothesis test using a p-value cut-off of 0.05, attempts to improve payouts for success, and results in greater rewards for countries with smaller reductions in incidence without a large increase in over-payment for countries where incidence had not declined. Using a higher less specific p-value cut-off, as in payout function 4, means that average payment for these countries rises further, but at the cost of a relatively high expected payment

when there is no real improvement. Payout functions 3 and 4 share with 1 and 2 the fact that the recipient would either receive the entire maximum payout award or nothing at all.

Instead of awarding either all or nothing, the payout can vary in amount with the level of significance reached, as in payout rules 5 and 6. These represent a compromise between using higher and lower p-value cut-off values.



**Figure 13:** Relationship between true reduction in incidence and the expected payout under a range of *threshold* payout functions (1-6) at a sample size of 20,000 (see Table 1). With each of these functions, there is scope for over and under-payment. For instance, at point A, where incidence actually increases but there is still a payout to the country, there is 'over-payment' on average; whilst at point B, where incidence declines but the country receives less than the full payout, there is 'under-payment' on average. The area to the left of the vertical dashed line is always the region of over-payment, but the region of under-payment is determined by the payout function being used.



**Figure 14:** Relationship between true reduction in incidence and the expected payout under a range of *continuous* payout functions (7-10) at a sample size of 20,000 (see Table 1). As in Figure 13 there are regions of average over and under-payment with respect to whether incidence is actually reduced. The area to left of the vertical dashed line is always the region of over-payment, but the region of under-payment is determined by the payout function being used.

Figure 14 shows the simulated relationship between the actual reduction in incidence and the expected payout profile for each of four continuous payout functions (function 7-10 in Table 1). With a linear increase in payout with measured reductions in incidence (function 7), the payout is typically lower than the payout on the basis of a classic hypothesis test (function 3). With a bonus associated with statistical significance (function 8), the expected payout is boosted for small and moderate incidence reductions with little increase in the expected overpayment when incidence did not really decrease. If the rate of payout is highly sensitive to small reductions in incidence (function 9) there is a substantial chance of over-payment and a diminished incentive for reductions in incidence greater than ~25 percent, since the expected payout rapidly reaches its maximum possible value. If the rate of payout is insensitive to small reductions in incidence but returns greater payouts for larger reductions in incidence (function 10), there is little risk of over-payment and no saturation at larger

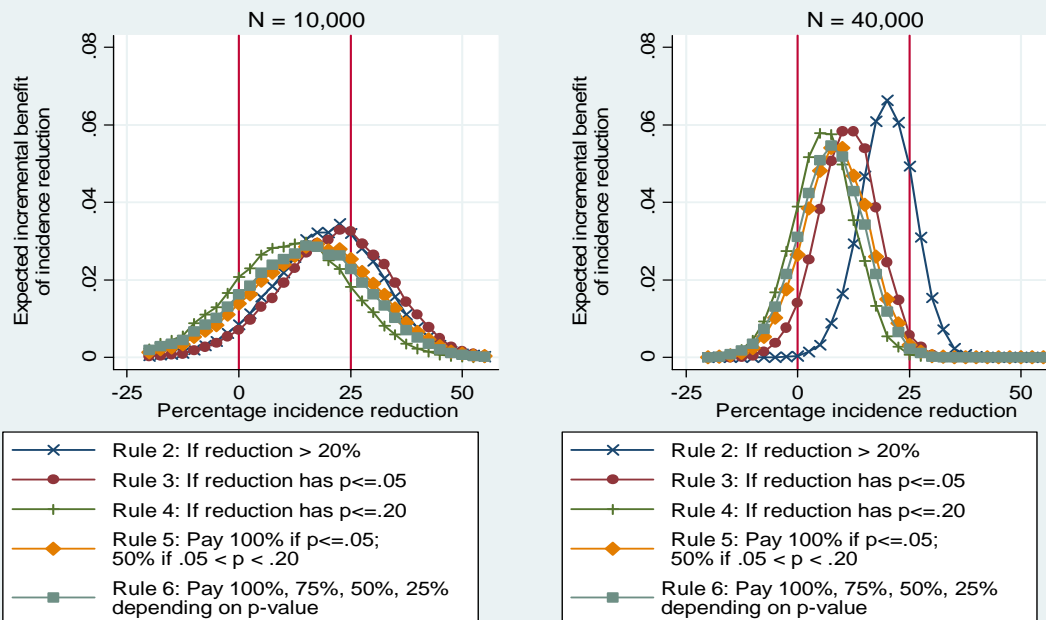


reductions. However, with this function, the expected payout, as a proportion of the maximum payout, overall is substantially reduced.

The choice of whether a payout function should be 'concave' or 'convex' (Figure 12 and functions 9 and 10) should be made carefully, as the effort-to-returns relationship in controlling infectious diseases is not linear, as it might be for another health intervention. First, program effort aimed at reducing the average number of new infections generated by each current infection (the basic reproductive rate,  $R_0$ , of the epidemic) is likely to suffer from diminishing returns, as the program is expanded beyond the early adopters to the increasingly recalcitrant. Second, the translation of a reduced basic reproductive rate into a reduction in prevalence/incidence depends on the baseline epidemiological context. Depending on the value of  $R_0$  and the epidemic context (shape of the curve in Figure 3), large reductions in incidence or prevalence could be generated by small changes in  $R_0$ , or small reductions could be generated by large changes. Further this relationship evolves over time and could differ by the type of intervention, and it is difficult to determine the nature of this relationship with the amount of information typically available in most of the countries with large HIV epidemics.

There are two reasons that donors will prefer the continuous payout functions to the threshold ones (1-6 in Table 1). First, when population survey data generates an estimate that HIV incidence has declined, when it really has not, the continuous functions produce only small expected payments to the country, whereas the threshold functions generate either the entire payment (for functions 1-2) or large fractions thereof. Second, the continuous payout functions offer a less concentrated expected incremental payout profile which is more likely to encourage the recipient to maximize incidence reduction, rather than only to aim for a target or, worse yet, to give up because the requirement for a payout seems unreachable.

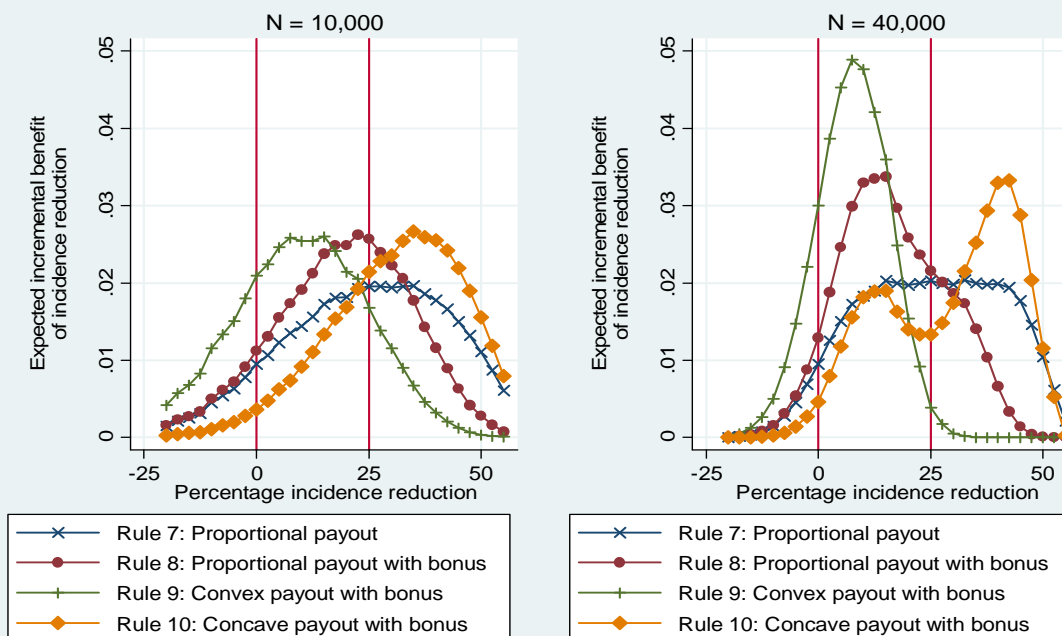
## Incremental impact of incidence reduction



Panel (a) Threshold functions 2 through 6, N= 10,000

Panel (b) Threshold functions 2 through 6, N= 40,000

## Incremental impact of incidence reduction



Panel (c) Continuous functions 7 through 10, N= 10,000

Panel (d) Continuous functions 7 through 10, N= 40,000

**Figure 15:** Incentive profiles for payout functions 2 through 10 at samples sizes 10,000 and 40,000 (see Table 1).

This second reason that donors might prefer continuous payout functions is illustrated in the four panels of Figure 15. The top two panels, panels (a) and (b), display the incentive profiles of the threshold payout functions 2 through 6, for sample sizes of 10,000 and 40,000, respectively. The bottom two panels show the incentives for the continuous payout functions 7 through 10, again for the two sample sizes. (The incentive profile of the first payout function is displayed for both sample sizes in Figure 11.) Comparing panels (a) and (c) for the smaller sample size with (b) and (d) with the larger one, we see that all of the threshold payout functions and the three continuous functions that include a bonus for statistical significance become more peaked, and therefore more concentrated on a limited range of incidence improvements, because increasing the sample makes them more discriminating between success and failure. If the donor's objective is to spread the incentive equally across a range of improvements from zero to fifty percent, this objective is best achieved with payout function 7 and a larger sample size.

Adding the bonus for statistical significance in function 8 increases the incentive for smaller improvement in comparison to function 7, but lowers it for larger improvements. Making the payout function convex as in function 9 looks like a bad idea, since it further concentrates the incentives on small improvements at the expense of reduced incentives for large improvements and an increased expected payout when improvement has not really occurred.

The last payout rule on the list, function 10, is a continuous concave function of measured improvement with a bonus for statistical significance. At the smaller sample size in panel (c), this payout function reduces the expected over-payment when there has been no true improvement while shifting most of the incentive to larger levels of true incidence reduction. The peak incentive for function 10 is at an incidence reduction of about 35 percent. At a sample size of 40,000, as depicted in panel (d), function 10's incentive profile resembles the back of a two-humped camel, with incentives concentrated first around a 12 percent reduction in incidence and then around a 40 percent reduction. This interesting profile might be agreed upon by the donor and recipient as an alternative to function 7 if they are not only unsure of how much incidence reduction can reasonably be achieved, but also wish to offer a much bigger expected payoff for each percentage reduction achieved beyond about 30 percent.

Once a payout function and sample size have been agreed to, the recipient country which aims to maximize its expected net benefits from the COD agreement would set the 'Marginal Expected Payout' equal to what it believes to be the 'Marginal Cost of the Effort' required to further reduce incidence. Suppose the incremental cost of decreasing incidence one more percentage point is constant and equal to 0.01 of the maximum award, or US\$10,000 if the maximum is one million

dollars.<sup>5</sup> Then the net COD-benefit-maximizing reduction in incidence can be read from each of the incentive profiles in Figure 15, by identifying the point where a descending marginal expected payout function intersects from above a horizontal line constructed at the value 0.01. For the functions and sample sizes in Figure 15 and assuming the value of the recipient's prevention effort per point of percentage reduction stays constant at one percent of the prize, the COD-benefit-maximizing reduction in incidence is between 30 and 45 percent in panel (a), between 20 and 30 percent in panel (b), between 30 and 55 percent in panel (c) and between 20 and 50 percent in panel (d)

This exploration of the incentive characteristics of the ten candidate payout functions suggests to us that donors will prefer a payout function to be continuous like function 7, possibly with either a bonus for a statistically significant improvement and as in function 8, or concavity as in function 10. As specified by Figure 12 and displayed in panels (c) and (d) of Figure 15, concavity limits the expected payout when incidence has not really declined and offers the highest incentives for the greatest achievement. A further advantage of payout function 10 is that, more than any other function considered here, it attenuates the incentive to the recipient to reduce the sample size in order to increase the chance of a payout even when no real improvement has occurred.

### **Corruption, Collusion, Capture, Competition and Unintended Consequences**

A frequent objection to the COD approach is that COD prize money, once awarded, might be stolen by government officials in the recipient country. In the words of Birdsall, Avedoff and Mahgoub:

COD Aid maximizes the recipient's discretion in using funds, making it virtually impossible for funders to ensure that COD Aid payments are used only for legitimate public purposes. But this is fundamentally true of all foreign aid. Even the most detailed monitoring of spending on traditional aid projects cannot guarantee that the recipient country is not taking advantage of the increased resources to release other funds for inappropriate uses<sup>6</sup>. The risk that COD Aid might encourage waste and corruption can be mitigated by establishing standards for public financial responsibility as a condition of eligibility [– the same same standards relied upon by

---

<sup>5</sup> Since the COD approach assumes that all costs of HIV prevention inputs and activities are financed from other donor or national sources, the cost function whose first derivative defines the COD-relevant marginal cost is the cost function for empowering and motivating the recipient country agents whose official or unofficial duties can influence the effectiveness of those already funded activities. With the promise of the COD payment, these agents might exert more effort in already planned and funded activities, might more successfully coordinate to achieve synergies among diverse activities, or might exercise their ingenuity to invent more effective activities than the central planners could ever have imagined or programmed.

<sup>6</sup> For example, a recent study finds that on average recipient countries adjust to each dollar increase in development assistance for health by reducing by a full dollar their tax-financed domestic spending for health (Lu et al. 2010).

the budget support assistance programs so popular with European donors and the World Bank]. Or, funders can accept the risk and rely on the improved assurance that, regardless of how the money is spent, progress has been achieved. [12]

Furthermore, some uses of the COD payout money which might be categorized as fraud by the inflexible rules of a specific donor, could instead be viewed as laudable examples of using aid resources to transparently motivate and reward key local contributors towards achieving the desired development objective, here a reduction in HIV incidence<sup>7</sup>. For example, earlier in this paper we suggested that the recipient government, as part of the launch of a COD for HIV prevention contract, might decide to publicize an offer to fund construction of a football stadium in the towns with the best performing municipal HIV prevention programs. For donors interested in furthering HIV prevention and aware of the impossibility of assuring that donor money never finances illegitimate expenses, this added flexibility of the COD approach will be an attractive feature.

To assure that COD aid rewards honestly measured HIV prevention, a COD program for HIV prevention must include strong safeguards against non-sampling sources of error, especially those that might be intentionally introduced by people who collude in hope of garnering some of the promised reward for themselves or their clients. In a recent study of the results-based assistance offered by the Global Alliance on Vaccines and Immunization (GAVI), the authors report that the estimates of vaccination achievement submitted to GAVI by recipient governments were frequently larger than the authors' own estimates of vaccination achievement derived from population surveys in the same country [91] The COD agreement signed between donor and recipient must specify sufficient safeguards to assure that the donor's proffered reward motivates people to work hard on HIV prevention, rather than on manipulating the measurement of HIV incidence.

The specific arrangements which will satisfy a given donor in a given context will vary with the strength of local statistical institutions and local governance. In some countries, the population surveys can be conducted by national firms with external audit. In others, where local statistical agencies either lack the internal auditing checks to assure data reliability or might be damaged by the unaccustomed pressure imposed by the COD arrangement, the surveys can be outsourced to an

---

<sup>7</sup> The economic optimality of some form of side payment is particularly apparent in the pursuit of a global public good like HIV prevention. Since no individual reaps more than a small fraction of the benefits of his or her HIV prevention enhancing efforts, financial and non-financial rewards serve to "internalize the positive externalities" of their socially desirable contributions. COD aid used to reward individual contributors to the COD objective, can be compared to conditional cash transfer programs which similarly internalize positive externalities. However, in contrast to the minutely planned, top-down rewards of a conditional cash transfer program, the COD approach leaves it to the recipient government's representatives to seek, perhaps through trial and error, a set of rewards which best advance the COD contract objective.

external firm. Whoever conducts the survey, its design can incorporate certain checks and balances to help control for systematic bias.

For example, the baseline and follow-up surveys could be conducted independently of the national surveillance efforts, minimising the chance of the selective inclusion/exclusion of clinic sites that generates an artificial impression of greater prevalence declines. As a further check, dummy variables could be created in the dataset that represent the dates of data collection/data entry and identifiers for the personnel doing the work. If the data has been collected and entered without any systematic error (include deliberate tampering), then the average rates of data collection and entry (forms per day) should be equal for all personnel, and the variables reflecting date and personnel should not be correlated with HIV prevalence. In addition, within-interviewer variation in HIV prevalence test results should be similar across all interviewers (an interviewer using her own blood for the tests would be detected by this test). A simple check could be made to verify this (e.g. logistic regression of HIV positive against the dummy variables), and the analysis could only be conducted subject to the data “passing” the test (i.e.  $p > 0.01$ , for no strong evidence of systematic error).

The advantages of the prevalence modelling method include its use of hard outcomes obtained from biological markers (indirectly HIV transmission patterns) rather than relying on reported sexual behaviour or programmatic activity [23]. The main data requirement – HIV test results in cross-section at the baseline and follow-up surveys – is straightforward to collect and the technology for reliable HIV testing is proven. Sample sizes for the baseline and follow-up surveys of 8,000-20,000 are usually adequate. The main disadvantage of this approach is the complexity of the process needed to interpret trends in prevalence, which could lead to disputes. The mathematical modelling method can be geared to be conservative, so that the threshold for finding evidence for behaviour change affecting the course of the epidemic is raised, by manipulating the model structure and parameter priors to maximise the natural declines in HIV prevalence. Alternatively, the model can be geared to be more generous by doing the opposite. But, the behavioural and epidemiological data are insufficiently precise to be sure that model priors are centered on ‘neutral’ territory (ie. *true* values of parameters). Another disadvantage of the prevalence modelling approach is the strong influence on the conclusions drawn of the information entered about HIV prevalence early in the epidemic before the surveillance system was set up. Informative prior distributions are usually derived from the subjective views of experts, perhaps supplemented by historical data recorded in the US Census Bureau International Population Database [92]. This makes the analysis vulnerable to gaming, unless the analysis is formally repeated under a range of assumptions about prior limits, or the expert opinion can be sourced reliably so that is, on the one hand, protected against capture by

recipient interests, and on the other perceived by all parties, including the recipient, to be fair and objective.

In some contexts it may be possible to engage the forces of competition in order to further stimulate HIV prevention effort. But safeguards must assure that introducing competition does not stimulate collusion and capture more than real HIV prevention effort. The principles described for finding evidence for incidence reductions are generic, so they can be applied to any country in the same manner. It follows that they could be applied to different regions of the same country. This may be an attractive option in large countries with semi-autonomous states, such as Nigeria, where the donor could be the federal government, perhaps financed by external partners, and the recipient the state government. In that case, the statistical power calculations and key results could be applied to each state of Nigeria, assuming independence between the states, and the payouts calculated separately. A concern for fairness would argue for adjusting each state's payout function so that a state with high prevention costs is rewarded for a less ambitious incidence reduction. However, adjusting the functions to the specifics of each state would be complex and vulnerable to charges of unfairness similar to those the procedure was intended to avoid. On the other hand, a concern to prevent the most Nigerian infections with a given aggregate COD award would argue for a single payout function, so that states with lower costs per HIV infection averted contribute correspondingly more averted HIV infections to the national total. Using a single payout function for every state would also have the advantages of simplicity and transparency, two criteria which might be more important than any others for the domestic and international acceptability of the COD approach.

If the surveys are organised at a national level, it would be important for them to afford the same statistical power to each state. This would likely incur extra costs, since a nationally-representative sample can use fewer sites (clusters) per state for the same overall statistical power. Alternatively, each state could organise its own survey, although this could incur a multiplication of fixed setup costs. In this situation, the payment to each state could be independent of those to other states, or states could compete for shares of a total reward budget. Competition for shares of a fixed maximum budget has the advantage of limiting a donor's total exposure and might achieve greater overall incidence reductions, due to greater effort by each state. However, to the degree that it reduces interstate cooperation for case detection and other key public health measures, competition might worsen overall national performance.<sup>8</sup>

---

<sup>8</sup> To save on the costs of powering many state-specific surveys, information could be shared across state boundaries. That is, state surveys of smaller samples sizes would be acceptable if the data between neighbouring states is 'pooled'. There are several mathematical methods for controlling the extent to which information is averaged among neighbours, the most common of which is Bayesian smoothing 93. Elliott P,

As with any pay-for-performance program, the COD approach might produce unintended consequences. For example, the existence of a COD contract for HIV prevention, with no comparable program for childhood immunization, might induce the recipient government to reallocate resources away from immunization and towards HIV prevention. Since the possibility that an aid program for one objective might compete with or crowd out unassisted programs with different objectives has always been a concern, it should not be an insuperable impediment to a COD for HIV prevention contract. Measures that can be taken to protect existing programs can include (1) maintenance of existing more traditional, input-financing, assistance to other programs; (2) keeping the maximum COD payment small in relation to the total HIV prevention budget; (3) including provisions regarding minimum performance of other health programs in the COD contract; and (4) reliance on the competence of the national ministry of health and/or the strength of the constituencies for other programs. Furthermore, since the COD approach does not specify how HIV prevention is to be achieved, it might stimulate more government expenditure on programs that the government deems to be important contributors to HIV prevention (e.g. road maintenance) than would a traditional input-oriented HIV prevention assistance package.

Since the earliest recognition that HIV is transmitted by sex and intravenous drug use, health experts have recognized that public health campaigns intended to stimulate public interest in preventing the spread of HIV infection might lead to unintended consequences, such as stigmatization or even persecution of those perceived to be members of groups likely to transmit [94]. To prevent this possibility, the United Nations Joint Programme on AIDS has published many guidelines on best practices for national AIDS programs to assure they do not stigmatize the HIV infected or those at most risk for contracting or transmitting the epidemic [95]. Some fear that a country that enters into a contract designed to reward a reduction in the number of new infections might, in its zeal to win the COD prize, carry out prevention programs which would infringe the human rights of its population.

The infringement of human rights for any reason is a cause for international concern. Just as the COD approach is unable to claim credit for HIV incidence reductions (due to the intentional absence of a rigorous evaluation design), the approach can never be shown to have caused any human rights infringements that occur during the life of the COD contract. Donors concerned about the possibility

---

Martuzzi M, Shaddick G (1995) Spatial statistical methods in environmental epidemiology: a critique. *Stat Methods Med Res* 4: 137-159. This method is typically used to produce estimates of incidence of a rare disease across small areas, where the sampling error (noise) would otherwise dominate. Furthermore, because smoothing tends to reduce the measured performance of the best performing states and increase that of the worst performing, its use would reduce the effect of a given state's effort on its payout, thereby undermining the objective of the COD approach.



that the COD contract they sponsor might be blamed for human rights abuses, or simply desirous of preventing such abuses, can impose eligibility requirements on recipient countries, such as human rights ombudsmen or civil-society watchdogs. Furthermore, they can include in the contract a provision that overtly discriminatory or stigmatizing behaviour towards the HIV-infected or the most at-risk populations, certified by a third-party, will annul the offer of the COD payment.

For any aid approach, each donor must weigh its desire to assist a recipient country against its concern that the country's governance structures are insufficient to support that kind of engagement. At one extreme, for some donors only traditional top-down, donor-executed paternalistic aid arrangements will be acceptable modes for helping some recipients. However, for donors and recipients able to engage in more mature relationships, such as those involved in recipient budget support, the COD approach offers an additional aid instrument which promises to improve HIV performance.

### **From policy discussion to pilot**

This paper demonstrates the statistical and computational feasibility of paying a country "cash-on-delivery" (COD) for an estimated reduction in HIV incidence achieved between two population-wide surveys. One of the two methods for estimating incidence reduction, the prevalence modelling approach, uses existing models and data and is ready to be piloted today. The other method, based on Tests for Recent Infection, is rapidly approaching readiness for field use, and might be usable within a COD agreement launched within the next year or two.<sup>9</sup> Use of either of these measurement approaches, in combination with one of the better performing payout functions analyzed in this paper, would enable a donor and recipient government to improve the motivation of the agents working on HIV prevention, perhaps sufficiently to dramatically reduce HIV incidence within a few years.

Some ask whether a donor's proposal of a COD arrangement to a recipient country implies that the donor doubts the sincerity of the recipient's desire to achieve the stated COD objective, here HIV prevention. But in view of the many alternative available sources of traditional foreign assistance, few of which demand any accountability for development results, a government AIDS program official who is insincere about HIV prevention could simplify his life by ignoring this aid modality

---

<sup>9</sup> The blood samples used to estimate baseline prevalence in the prevalence modelling approach could be conserved. If the methods for the test of recent infection approach become available before the follow-up survey has been completed, the TRI approach can be applied to samples from both surveys and its results compared to those of the prevalence modelling approach on the same population.

altogether. To the contrary, it seems likely that the recipient government official who will be most interested in a COD-for-HIV-prevention contract is one who is intensely interested in improving HIV prevention, aware that effective prevention requires the concerted engagement of many actors at all levels of government and civil-society and frustrated by his inability to get the devoted attention of these other actors. Such a visionary leader will see a COD contract as an opportunity to engage these others by offering them pieces of the COD payout, either in kind (such as the football stadiums mentioned above) or in the form of financial support for their other underfunded priorities.

While space constraints have limited our treatment of the political economy arguments in favour of a COD approach to foreign assistance in general and HIV prevention in particular, we are aware that more discussion of such arguments for and against the COD approach would be required before such a novel mechanism could be deployed for HIV prevention. Research to strengthen the technical foundation of COD will contribute to the policy debate to the extent that it relieves concerns about the approach's vulnerability to the choice of the payout function or to sampling or non-sampling errors. Among non-sampling errors, donors and the international public health community, including its members and other concerned citizens within potential recipient countries, will be most concerned that the COD approach could be corrupted through collusion among the measurement agency staff or through capture of the body of experts that will approve the estimates of HIV incidence that determine the size of the COD payout.

Further research on payout functions could precisely define the concept of an "optimal payout function" and demonstrate how that optimum could be identified in a specific epidemiological and health systems context. Assuming that the donor's objective is to motivate the recipient to achieve the maximum incidence reduction per dollar of COD reward paid, the optimal payout function from the donor's perspective will depend upon an ex ante assessment of how difficult it will be for the recipient to achieve any given incidence reduction over a given number of years between the two surveys. Formally, this assessment amounts to a judgement or econometric estimate of the shape of the recipient's incremental cost function for HIV incidence reduction.<sup>10</sup> Research could demonstrate the derivation of an optimal payout function in a specific real-world country, though its empirical validation would have to await that function's deployment as part of a real COD program in that country.

Vulnerability of the COD to sampling errors depends on the sample size, the survey interval and the choice of the payout function. By choosing a continuous rather than a threshold payout function, the COD program aligns the interests of the recipient more closely with those of the donor in favour

---

<sup>10</sup> Note 5 discusses the concept of "incremental costs of HIV incidence reduction" in the COD context.

of achieving the largest possible reduction in HIV incidence, rather than focusing on a target reduction. Though the COD donor will typically set an approximate maximum budget for the baseline and follow-up surveys, the number of individuals which can be surveyed for that budget will depend on the unit cost per individual. The recipient country can help to assure a large sample size by encouraging the most efficient local survey firms to bid their lowest prices on the survey contract. Further work on the COD can develop a model bidding document which facilitates efficient sampling by requesting prospective bidders to propose a sample size for a given budget rather than the more common approach which is to propose a budget for a given sampling problem.

Further research can also explore the degree to which repeating the COD agreement several times in the same population, with the follow-up survey from the first agreement serving as the baseline for the second agreement, can reduce the vulnerability of the COD to sampling and non-sampling errors. For example, it may be the case that an under-estimate of incidence reduction and the associated underpayment at the end of a first agreement would be offset on average by overpayment on the second agreement, and vice-versa.

Further research would also be useful on ways to design and manage a survey in order to minimize non-sampling errors, including collusion among survey or information management staff or capture of the technical experts for the purpose of defrauding the COD agreement. For example, randomized re-interviewing of already sampled individuals might be important for dissuading those who might try to cheat. In situations where multiple states are competing for shares of a single COD reward budget, perhaps state survey staff could be randomized to states other than their own.

While the COD approach intentionally avoids the impact evaluation paradigm, the approach could be evaluated in a given country context if the follow-up and baseline surveys and the incidence measurement approach were replicated in several regions of a country, only some of which were randomly assigned a COD agreement with a payout function. We hesitate to propose such an evaluation because of the large sample sizes and/or long survey intervals that would be needed to find statistically significant differences in HIV cases averted between the two arms of such an evaluation. Instead we believe that, in a country with a serious HIV/AIDS epidemic, the successful implementation of a COD for HIV prevention program complete with payouts being awarded for incidence reduction would be sufficient evidence of the program's worth.

If the idea of using a "cash-on-delivery" reward for HIV incidence reduction catches the imagination of one or more donors and if countries where many of the HIV prevention input and activity budgets are already financed are interested in reaching for this prize, then we believe there would be much

to gain from piloting this idea in a country, or perhaps in a region of a county. The COD approach to prevention could also be piloted with another disease, such as malaria, where the challenge of measuring averted cases or of attributing them to specific interventions is also an impediment to results-based public health programming. Just as in candidate sectors outside of public health, wherever there is awareness that the agents of program success lack motivation to exert sufficient coordinated productive effort towards the program's goals, the cash-on-delivery approach holds the promise of enhancing the effectiveness of government programs and of the existing bilateral and multilateral aid that supports them.

## References

1. Kevin Palmer (personal communication) (2010).
2. Committee for the Evaluation of the President's Emergency Plan for AIDS Relief (PEPFAR) Implementation, Sepulveda J, Carpenter C, Curran J, Holzemer W, et al. (2007) PEPFAR Implementation: Progress and Promise. Washington, D.C.: Institute of Medicine of the National Academies.
3. The Global Fund to Fight AIDS Tuberculosis and Malaria The Global Fund to Fight AIDS, Tuberculosis and Malaria. [www.theglobalfund.org](http://www.theglobalfund.org).
4. Over AM (2008) Opportunities for presidential leadership on AIDS: From an emergency program to a sustainable policy. In: Birdsall N, editor. *White House and the World: A global development agenda for the next US President*. Washington, D.C.: Brookings Institution Press.
5. Over AM (2008) Prevention Failure: The Ballooning Entitlement Burden of U.S. Global AIDS Treatment Spending and What to Do About It. Washington, D.C.: Center for Global Development.
6. Stover J, Walker N, Garnett GP, Salomon JA, Stanecki KA, et al. (2002) Can we reverse the HIV/AIDS pandemic with an expanded response? *Lancet* 360: 73-77.
7. Ainsworth M, Over AM (1997) *Confronting AIDS: public priorities in a global epidemic*. New York: Oxford University Press.
8. Over M, Piot P (1996) Human immunodeficiency virus infection and other sexually transmitted diseases in developing countries: public health importance and priorities for resource allocation. *J Infect Dis* 174: S162-175.
9. Kaneko A A community-directed strategy for sustainable malaria elimination on islands: short-term MDA integrated with ITNs and robust surveillance. *Acta Trop* 114: 177-183.
10. Atkinson JA, Fitzgerald L, Toaliu H, Taleo G, Tynan A, et al. Community participation for malaria elimination in Tafea Province, Vanuatu: Part I. Maintaining motivation for prevention practices in the context of disappearing disease. *Malar J* 9: 93.
11. Barder O, Birdsall N (2006) *Payments for Progress: A Hands-Off Approach to Foreign Aid*. [Center for Global Development: Working Paper 102].
12. Birdsall N, Savedoff W, Mahgoub A, Vyborny K (2010) *Cash on Delivery: A New Approach to Foreign Aid, with an Application to Primary Schooling*. Baltimore, MD, USA: Brookings Institution Press.
13. Radelet S (2003) *A Policymaker's Guide to the Millenium Challenge Account*. Washington, D.C.: Center for Global Development.
14. Eichler R, Levine R (2009) *Performance-Based Incentives Working Group. 2009. Performance Incentives for Global Health: Potential and Pitfalls*. Baltimore, MD, USA: Brookings Institution Press.
15. Easterly W (2006) *The White Man's Burden: Why the West's Efforts to Aid Africa Have Done So Much Ill and So little good*. New York: Penguin.
16. Filmer D, Pritchett L (1999) The impact of public spending on health: does money matter? *Soc Sci Med* 49: 1309-1323.
17. Guinness L, Kumaranayake L, Rajaraman B, Sankaranarayanan G, Vannela G, et al. (2005) Does scale matter? The costs of HIV-prevention interventions for commercial sex workers in India. *Bull World Health Organ* 83: 747-755.
18. Dandona L, Sisodia P, Prasad TL, Marseille E, Chalapathi Rao M, et al. (2005) Cost and efficiency of public sector sexually transmitted infection clinics in Andhra Pradesh, India. *BMC Health Serv Res* 5: 69.
19. Marseille E, Dandona L, Marshall N, Gaist P, Bautista-Arredondo S, et al. (2007) HIV prevention costs and program scale: data from the PANCEA project in five low and middle-income countries. *BMC Health Serv Res* 7: 108.

20. Bautista-Arredondo S, Gadsden P, Harris JE, Bertozzi SM (2008) Optimizing resource allocation for HIV/AIDS prevention programmes: an analytical framework. *AIDS* 22 Suppl 1: S67-74.
21. Merson M, Padian N, Coates TJ, Gupta GR, Bertozzi SM, et al. (2008) Combination HIV prevention. *Lancet* 372: 1805-1806.
22. Halperin DT (2009) Combination HIV prevention must be based on evidence. *Lancet* 373: 544-545.
23. Hallett TB, White PJ, Garnett GP (2007) Appropriate evaluation of HIV prevention interventions: from experiment to full-scale implementation. *Sex Transm Infect* 83: i55-60.
24. Sherr L, Lopman B, Kakowa M, Dube S, Chawira G, et al. (2007) Voluntary counselling and testing: uptake, impact on sexual behaviour, and HIV incidence in a rural Zimbabwean cohort. *AIDS* 21: 851-860.
25. Corbett EL, Makamure B, Cheung YB, Dauya E, Matambo R, et al. (2007) HIV incidence during a cluster-randomized trial of two strategies providing voluntary counselling and testing at the workplace, Zimbabwe. *AIDS* 21: 483-489.
26. Hallett TB, Dube S, Cremin I, Lopman B, Mahomva A, et al. (2009) The Role of Testing and Counselling For HIV Prevention and Care in the Era of Scaling-Up Antiretroviral Therapy. *Epidemics* 1: 77-82.
27. Gregson S, Zhuwau T, Ndlovu J, Nyamukapa CA (2002) Methods to reduce social desirability bias in sex surveys in low-development settings: experience in Zimbabwe. *Sex Transm Dis* 29: 568-575.
28. Garnett GP, Anderson RM (1993) Factors controlling the spread of HIV in heterosexual communities in developing countries: patterns of mixing between different age and sexual activity classes. *Philos Trans R Soc Lond B Biol Sci* 342: 137-159.
29. Gregson S, Nyamukapa CA, Garnett GP, Mason PR, Zhuwau T, et al. (2002) Sexual mixing patterns and sex-differentials in teenage exposure to HIV infection in rural Zimbabwe. *Lancet* 359: 1896-1903.
30. Bracher M, Santow G, Watkins SC (2004) Assessing the potential of condom use to prevent the spread of HIV: a microsimulation study. *Stud Fam Plann* 35: 48-64.
31. Hallett TB, Aberle-Grasse J, Bello G, Boulos LM, Cayemittes MPA, et al. (2006) Declines in HIV prevalence can be associated with changing sexual behaviour in Uganda, urban Kenya, Zimbabwe, and urban Haiti. *Sex Transm Infect* 82: i1-8.
32. Morris M, Kretzschmar M (1997) Concurrent partnerships and the spread of HIV. *AIDS* 11: 641-648.
33. Hallett TB, Gregson S, Lewis JJC, Lopman BA, Garnett GP (2007) Behaviour change in generalised HIV epidemics: impact of reducing cross-generational sex and delaying age at sexual debut. *Sex Transm Infect* 83: i50-54.
34. Anderson RM, May RM, Boily MC, Garnett GP, Rowley JT (1991) The spread of HIV-1 in Africa: sexual contact patterns and the predicted demographic impact of AIDS. *Nature* 352: 581-589.
35. UNAIDS Reference Group on Estimates Modelling and Projections (2002) Improved methods and assumptions for estimation of the HIV/AIDS epidemic and its impact: Recommendations of the UNAIDS Reference Group on Estimates, Modelling and Projections. *AIDS* 16: W1-14.
36. Kilian AH, Gregson S, Ndyabangi B, Walusaga K, Kipp W, et al. (1999) Reductions in risk behaviour provide the most consistent explanation for declining HIV-1 prevalence in Uganda. *AIDS* 13: 391-398.
37. Hallett TB, Gregson S, Gonese E, Mugurungi O, Garnett GP (2009) Assessing evidence for behaviour change affecting the course of HIV epidemics: A new mathematical modelling approach and application to data from Zimbabwe *Epidemics* 1: 108-117 (doi:110.1016/j.epidem.2009.1003.1001).

38. Todd J, Glynn JR, Marston M, Lutalo T, Biraro S, et al. (2007) Time from HIV seroconversion to death: a collaborative analysis of eight studies in six low and middle-income countries before highly active antiretroviral therapy. *AIDS* 21: S55-S63.
39. Boily MC, Godin G, Hogben M, Sherr L, Bastos FI (2005) The impact of the transmission dynamics of the HIV/AIDS epidemic on sexual behaviour: a new hypothesis to explain recent increases in risk taking-behaviour among men who have sex with men. *Med Hypotheses* 65: 215-226.
40. WHO, UNAIDS, UNICEF (2008) Towards Universal Access: Scaling up priority HIV/AIDS interventions in the health sector Progress Report (available from [http://www.who.int/hiv/pub/towards\\_universal\\_access\\_report\\_2008.pdf](http://www.who.int/hiv/pub/towards_universal_access_report_2008.pdf)). Geneva.
41. Braitstein P, Brinkhof MW, Dabis F, Schechter M, Boule A, et al. (2006) Mortality of HIV-1-infected patients in the first year of antiretroviral therapy: comparison between low-income and high-income countries. *Lancet* 367: 817-824.
42. Salomon JA, Hogan DR (2008) Evaluating the impact of antiretroviral therapy on HIV transmission. *AIDS* 22 Suppl 1: S149-159.
43. Lopman BA, Nyamukapa C, Hallett TB, Mushati P, Spark-du Preez N, et al. (2009) Role of widows in the heterosexual transmission of HIV in Manicaland, Zimbabwe, 1998-2003. *Sex Transm Infect* 85 Suppl 1: i41-48.
44. Marston M, Harriss K, Slaymaker E (2008) Non-response bias in estimates of HIV prevalence due to the mobility of absentees in national population-based surveys: a study of nine national surveys. *Sex Transm Infect* 84 Suppl 1: i71-i77.
45. Reniers G, Eaton J (2009) Refusal bias in HIV prevalence estimates from nationally representative seroprevalence surveys. *AIDS* 23: 621-629.
46. Brookmeyer R (2009) Should biomarker estimates of HIV incidence be adjusted? *AIDS* 23: 485-491.
47. Garnett GP, Gregson S (2000) Monitoring the course of the HIV-1 epidemic: the influence of patterns of fertility on HIV-1 prevalence estimates. *Mathematical Population Studies* 8: 251-277.
48. Grassly NC, Lowndes CM, Rhodes T, Judd A, Renton A, et al. (2003) Modelling emerging HIV epidemics: the role of injecting drug use and sexual transmission in the Russian Federation, China and India. *International Journal of Drug Policy* 14: 25-43.
49. Poole D, Raftery AE (2000) Inference for deterministic simulation models: The Bayesian melding approach. *Journal of the American Statistical Association* 95: 1244-1255.
50. Powers KA, Poole C, Pettifor AE, Cohen MS (2008) Rethinking the heterosexual infectivity of HIV-1: a systematic review and meta-analysis. *Lancet Infect Dis*.
51. Boily MC, Baggaley RF, Wang L, Masse B, White RG, et al. (2009) Heterosexual risk of HIV-1 infection per sexual act: systematic review and meta-analysis of observational studies. *Lancet Infect Dis* 9: 118-129.
52. Garnett GP, Hughes JP, Anderson RM, Stoner BP, Aral SO, et al. (1996) Sexual mixing patterns of patients attending sexually transmitted diseases clinics. *Sex Transm Dis* 23: 248-257.
53. Alkema L, Raftery AE, Brown T (2008) Bayesian melding for estimating uncertainty in national HIV prevalence estimates. *Sex Transm Infect* 84 Suppl 1: i11-i16.
54. Alkema L, Raftery AE, Clark SJ (2007) Probabilistic projections of HIV prevalence using Bayesian melding. *The Annals of Applied Statistics*. pp. 229-248.
55. Ghys PD, Kufa E, George MV, for the Unaid Reference Group on Estimates Modelling and Projections' ad hoc Working Group on interpreting trends in prevalence and incidence of HIV infection in countries with generalised epidemics (2006) Measuring trends in prevalence and incidence of HIV infection in countries with generalised epidemics. *Sex Transm Infect* 82: i52-56.
56. Laeyendecker O, Rothman RE, Henson C, Horne BJ, Ketlogetswe KS, et al. (2008) The effect of viral suppression on cross-sectional incidence testing in the Johns Hopkins hospital emergency department. *J Acquir Immune Defic Syndr* 48: 211-215.

57. Shisana O, Rehle T, Simbayi L, Zuma K, Jooste S, et al. (2009) South African National Prevalence, Incidence, Behaviour and Communication Survey, 2008. A Turning Tide Among Teenagers? Cape Town, South Africa: HSRC Press
58. National AIDS and STI Control Programme, Ministry of Health - Kenya (2008) Kenya AIDS Indicator Survey, 2007: Preliminary Report; KAIS presentation, 7th October, George Washington University, Washington D.C., USA.
59. Wawer MJ, Gray RH, Sewankambo NK, Serwadda D, Li X, et al. (2005) Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda. *J Infect Dis* 191: 1403-1409.
60. Hollingsworth TD, Anderson RM, Fraser C (2008) HIV-1 transmission, by stage of infection. *J Infect Dis* 198: 687-693.
61. Gouws E, Mishra V, Fowler TB (2008) Comparison of adult HIV prevalence from national population-based surveys and antenatal clinic surveillance in countries with generalised epidemics: implications for calibrating surveillance data. *Sex Transm Infect* 84 Suppl 1: i17-i23.
62. Lewis JJ, Ronsmans C, Ezeh A, Gregson S (2004) The population impact of HIV on fertility in sub-Saharan Africa. *AIDS* 18 Suppl 2: S35-43.
63. Jeffreys H (1961) The theory of probability. New York, USA: Oxford University Press.
64. Raftery A (1996) Hypothesis testing and model selection. . In: Gilks W, Richardson S, Spiegelhalter D, editors. *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall.
65. Boerma JT, Ghys PD, Walker N (2003) Estimates of HIV-1 prevalence from national population-based surveys as a new gold standard. *Lancet* 362: 1929-1931.
66. Zaba B, Boerma T, White R (2000) Monitoring the AIDS epidemic using HIV prevalence data among young women attending antenatal clinics: prospects and problems. *AIDS* 14: 1633-1645.
67. Low-Beer D, Afkhami H, Komatsu R, Banati P, Sempala M, et al. (2007) Making performance-based funding work for health. *PLoS Med* 4: e219.
68. Hallett TB, Zaba B, Todd J, Lopman B, Mwita W, et al. (2008) Estimating incidence from prevalence in generalised HIV epidemics: methods and validation. *PLoS Med* 5: e80.
69. Stover J, Johnson P, Zaba B, Zwahlen M, Dabis F, et al. (2008) The Spectrum projection package: improvements in estimating mortality, ART needs, PMTCT impact and uncertainty bounds. *Sex Transm Infect* 84 Suppl 1: i24-i30.
70. McDougal JS, Pilcher CD, Parekh BS, Gershy-Damet G, Branson BM, et al. (2005) Surveillance for HIV-1 incidence using tests for recent infection in resource-constrained countries. *AIDS* 19 Suppl 2: S25-30.
71. Murphy G, Parry JV (2008) Assays for the detection of recent infections with Human Immunodeficiency Virus Type 1. *Eurosurveillance* 13: 4-10.
72. Parekh BS, Kennedy MS, Dobbs T, Pau CP, Byers R, et al. (2002) Quantitative detection of increasing HIV type 1 antibodies after seroconversion: a simple assay for detecting recent HIV infection and estimating incidence. *AIDS Res Hum Retroviruses* 18: 295-307.
73. McDougal JS, Parekh BS, Peterson ML, Branson BM, Dobbs T, et al. (2006) Comparison of HIV type 1 incidence observed during longitudinal follow-up with incidence estimated by cross-sectional analysis using the BED capture enzyme immunoassay. *AIDS Res Hum Retroviruses* 22: 945-952.
74. Barnighausen T, Wallrauch C, Welte A, McWalter TA, Mbizana N, et al. (In Press, 2008) HIV incidence in rural South Africa: comparison of estimates from longitudinal surveillance and cross-sectional cBED assay testing. *PLoS ONE*.
75. Barnighausen T, Rosner Z, Welte A, McWalter TA (In preparation) Worldwide application of the capture of BED immunoassay to estimate HIV incidence: a systematic review.
76. Karita E, Price M, Hunter E, Chomba E, Allen S, et al. (2007) Investigating the utility of the HIV-1 BED capture enzyme immunoassay using cross-sectional and longitudinal seroconverter specimens from Africa. *AIDS* 21: 403-408.



77. Hargrove JW, Humphrey JH, Mutasa K, Parekh BS, McDougal JS, et al. (2008) Improved HIV-1 incidence estimates using the BED capture enzyme immunoassay. *AIDS* 22: 511-518.
78. UNAIDS Reference Group on Estimates Modelling and Projections (2006) Statement on the use of the BED-assay for the estimation of HIV-1 incidence for surveillance or epidemic monitoring. *Wkly Epidemiol Rec* 81: 40-41.
79. McWalter TA, Welte A (2009) Relating recent infection prevalence to incidence with a sub-population of assay non-progressors. *J Math Biol*.
80. Surveillance and Survey and the Laboratory Working Groups to the Office of the Global AIDS Coordinator (2006) Interim Recommendations For The Use Of The BED Capture Enzyme Immunoassay For Incidence Estimation and Surveillance. (Available from: [http://www.cdc.gov/nchstp/od/GAP/docs/surveillance/Interim%20Recommendations%20for%20the%20use%20of%20the%20BED%20capture%20enzyme%20immunoassay%20for%20incidence%20estimation%20and%20surveillance%20Approved%20November%2021%202006%20\(2\).pdf](http://www.cdc.gov/nchstp/od/GAP/docs/surveillance/Interim%20Recommendations%20for%20the%20use%20of%20the%20BED%20capture%20enzyme%20immunoassay%20for%20incidence%20estimation%20and%20surveillance%20Approved%20November%2021%202006%20(2).pdf)).
81. Kim A, McDougal S, Hargrove J, Morgan M, Nolan M, et al. (2007) Towards more plausible estimates of HIV incidence in cross-sectional serologic surveys in Africa? Application of a HIV-1 incidence assay with post-assay adjustment. [abstract V-190]. Fourteenth Conference on Retroviruses and Opportunistic Infections. Los Angeles, USA.
82. Barnighausen T, Wallrauch C, Welte A, McWalter TA, Mbizana N, et al. (2008) HIV incidence in rural South Africa: comparison of estimates from longitudinal surveillance and cross-sectional cBED assay testing. *PLoS ONE* 3: e3640.
83. Laeyendecker O. Session II: Measuring new HIV infections; where we are with new technologies and approaches measuring HIV incidence. BED + avidity testing algorithm; 2009; Bangkok, Thailand.
84. Hallett TB, Ghys P, Barnighausen T, Yan P, Garnett GP (2009) Errors in 'BED'-derived estimates of HIV incidence will vary by place, time and age. *PLoS ONE* 4: e5720.
85. Marinda ET, Hargrove JW, Slabbert H, van Zyl G, Levin J, et al. (Submitted, 2009) Performance of the BED Capture Enzyme Immunoassay Among HIV-1 Patients Immediately Prior to, and During, Anti-retroviral Therapy. *AIDS*.
86. Hladik W, Olara D, Were W, Mermin J, Downing R. The Effect of Antiretroviral Treatment on the Specificity of the Serological BED HIV-1 Incidence Assay (Abstract 998); 2007 June, 2007; Kigali, Rwanda.
87. Hayashida T, Gatanaga H, Tanuma J, Oka S (2008) Effects of low HIV type 1 load and antiretroviral treatment on IgG-capture BED-enzyme immunoassay. *AIDS Res Hum Retroviruses* 24: 495-498.
88. Hubert JB, Burgard M, Dussaix E, Tamalet C, Deveau C, et al. (2000) Natural history of serum HIV-1 RNA levels in 330 patients with a known date of infection. The SEROCO Study Group. *AIDS* 14: 123-131.
89. Welte T, McWalter TA, Laeyendecker O, Hallett TB (2009) Using Tests for Recent Infection to Estimate Incidence: Problems and Prospects for HIV. *Eurosurveillance* Submitted.
90. Welte A, McWalter TA, Barnighausen T (2009) Response to 'Should biomarker estimates of HIV incidence be adjusted?' *AIDS* In Press.
91. Lu C, Michaud CM, Gakidou E, Khan K, Murray CJ (2006) Effect of the Global Alliance for Vaccines and Immunisation on diphtheria, tetanus, and pertussis vaccine coverage: an independent assessment. *Lancet* 368: 1088-1095.
92. U.S. Census Bureau International Population Database (2007) Midyear Population, by Age and Sex. available from <http://www.census.gov/ipc/www/worldhtml>.
93. Elliott P, Martuzzi M, Shaddick G (1995) Spatial statistical methods in environmental epidemiology: a critique. *Stat Methods Med Res* 4: 137-159.
94. Tarantola D, Mann J (1995) AIDS and human rights. *AIDS Soc* 6: 1, 5.

95. UNAIDS (1997) Reducing HIV Stigma and Discrimination: a critical part of national AIDS programmes. Geneva.
96. Marston M, Todd J, Glynn JR, Nelson K, Rangsin R, et al. (2007) Estimating 'net' HIV-related mortality and the importance of background mortality rates. *AIDS* 21: S65-S71.
97. Walker PT, Hallett TB, White PJ, Garnett GP (2008) Interpreting declines in HIV prevalence: impact of spatial aggregation and migration on expected declines in prevalence. *Sex Transm Infect* 84 Suppl 2: ii42-48.
98. Schneider M, Zwahlen M, Egger M (2004) Natural history and mortality in HIV-positive individuals living in resource-poor settings. Available from [www.epidem.org](http://www.epidem.org).
99. Hallett TB, Gregson S, Dube S, Garnett GP (2008) The Impact of Monitoring HIV Patients Prior to Treatment in Resource-Poor Settings: Insights from Mathematical Modelling. *PLoS Med* 5: e53.
100. Cheluget B, Baltazar G, Orege P, Ibrahim M, Marum LH, et al. (2006) Evidence for population level declines in adult HIV prevalence in Kenya. *Sex Transm Infect* 82 Suppl 1: i21-26.
101. Padian NS, Shiboski SC, Jewell NP (1991) Female-to-male transmission of human immunodeficiency virus. *JAMA* 266: 1664-1667.
102. Nicolosi A, Correa Leite ML, Musicco M, Arici C, Gavazzeni G, et al. (1994) The efficiency of male-to-female and female-to-male sexual transmission of the human immunodeficiency virus: a study of 730 stable couples. Italian Study Group on HIV Heterosexual Transmission. *Epidemiology* 5: 570-575.
103. World Health Organization, UNAIDS, UNICEF (2007) Towards Universal Access: Scaling up priority HIV/AIDS interventions in the health sector. [http://www.who.int/hiv/mediacentre/universal\\_access\\_progress\\_report\\_enpdf](http://www.who.int/hiv/mediacentre/universal_access_progress_report_enpdf) (accessed 30/10/07).
104. Gregson S, Garnett GP, Nyamukapa CA, Hallett TB, Lewis JJ, et al. (2006) HIV decline associated with behavior change in eastern Zimbabwe. *Science* 311: 664-666.
105. Rottingen JA, Cameron DW, Garnett GP (2001) A systematic review of the epidemiologic interactions between classic sexually transmitted diseases and HIV: how much really is known? *Sex Transm Dis* 28: 579-597.

## Supplementary Appendix 1. Details Of The Simulation Model Used To Generate Counter-Factual Projections.

(Based on the methods in Hallett *et al.* Epidemics, 2009)

Following others [31,36,47], the model represents the heterosexual spread of HIV in a sex and sexual activity stratified population. The model is described in the following four sections: (i) Differential equations; (ii) Calculating the force of infection; (iii) Changes in sexual behaviour parameters; (iv) Model outputs; and, (v) Fixed parameter values.

### (i) Differential equations

The model is defined by a set of ordinary differential equations which are solved numerically using custom-made software developed by the authors. The state variables are given by  $X_{k,l}^s$ :  $s$  is the infection-status (1= susceptible; 2= acute infection; 3,4,5= latent infection; 6= pre-AIDS; 7= AIDS; 8=pre-AIDS and will receive antiretroviral therapy (ART), 9= on ART),  $k$  is sex (1= female; 2= male) and  $l$  is the sexual activity risk group (1= (highest risk), 2 and 3= (lowest risk)). Sexual activity group 1 represents those with high numbers of casual partners; group 2 represents those with long-term casual partners; and, 3 represents those with in stable spousal partnerships. Groups 1 and 2 together are classified as the 'higher risk groups'. In this document, the  $\bullet$  symbol indicates the sum of the state variable across that dimension, and  $X$  (without subscripts) is the sum across all strata (equivalent to  $X_{\bullet,\bullet}$ ).

The ordinary differential equations describing changes in the state variable over time are as follows, and the course of HIV infection is illustrated in Figure 1 in the main text.

Latent infection is split into three compartments so that the overall distribution of survival from HIV infection to death in the absence of treatment approximates a gamma distribution, which is indicated by the available data from resource-poor settings (Figure S1).

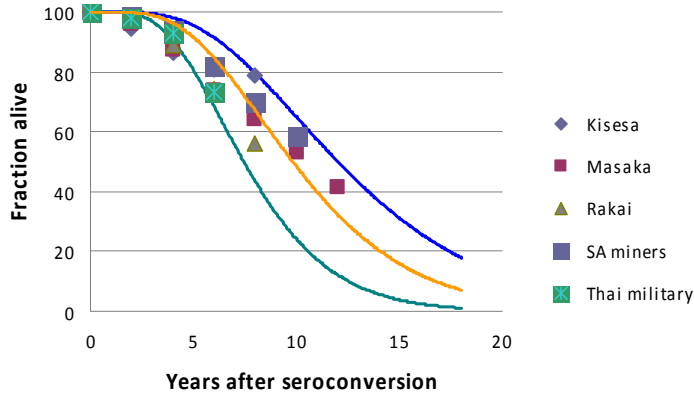
$$\begin{aligned} \frac{dX_{k,l}^1}{dt} &= \frac{(\mu + \alpha)}{2} \phi_{k,l} X_{k,l}^{\bullet} - \lambda_{1,l} X_{k,l}^1 - \mu X_{k,l}^1 \\ \frac{dX_{k,l}^2}{dt} &= \lambda_{k,l} X_{k,l}^1 - (\mu + \sigma_2) X_{k,l}^2 \\ \frac{dX_{k,l}^s}{dt} &= \sigma_{s-1} X_{k,l}^{s-1} - (\mu + \sigma_s) X_{k,l}^s \quad \} s = 3,4,5 \\ \frac{dX_{k,l}^6}{dt} &= (1 - a(t)) \sigma_5 X_{k,l}^5 - (\mu + \sigma_6) X_{k,l}^6 \\ \frac{dX_{k,l}^7}{dt} &= \sigma_6 X_{k,l}^6 - (\mu + \sigma_7) X_{k,l}^7 \\ \frac{dX_{k,l}^8}{dt} &= a(t) \sigma_5 X_{k,l}^5 - (\mu + \sigma_8) X_{k,l}^8 \\ \frac{dX_{k,l}^9}{dt} &= \sigma_8 X_{k,l}^8 - (\mu + \sigma_9) X_{k,l}^9 \end{aligned}$$

Here, the rate of individuals ceasing sexual activity is  $\mu$  (mean duration sexually active  $1/\mu$ ); in the absence of AIDS, the population grows exponentially at a rate  $\alpha$ ; the fraction of men and women starting sexual activity in each sexual activity group is  $\phi_{k,l}$  (such that  $\phi_{k,\bullet} = 1$ ); the force of infection for individuals in each sex and sexual activity group is  $\lambda_{k,l}$ , which is defined below.

The fraction of individuals progressing to pre-AIDS that will start treatment is given by  $a(t)$ , and this changes over time in the following way:

$$a(t) = \max[0, \min(a_{\max}, a_{\text{rate}}(t - a_{\text{start}}))] ]$$

where,  $a_{\max}$  is the maximum level of coverage achieved,  $a_{\text{start}}$  is the time in the simulation when ART coverage starts to increase and  $a_{\text{rate}}$  is the increase in coverage.



**Figure S1:** Net survival distribution with HIV. The yellow line show the best fit of the gamma distribution to the observational data, reported by Marston et al.[96]. The green and blue lines, respectively, show the limits of the prior distribution for mean survival with HIV infection, in the absence of ART.

In the model, it can be assumed that individuals do not change sexual activity group during their lifetime and that the proportion of individuals entering each group is constant over time (this assumption is labelled ‘no replacement’). Over the course of HIV epidemics, some sexual activity groups may suffer greater AIDS-related mortality than others, leading to changes in the overall distribution of risk in the population. The model can counteract that change and allow individuals to move between groups in such a way that the fraction of adult men and women in each risk-group remains constant over time (this assumption is labelled ‘with replacement’). The replacement of individuals in the higher risk groups is simulated in the following way:

for  $l = 1, 2$ :

$$G_{k,l} = \phi_{k,l} X_{k,\bullet}^{\bullet} - X_{k,l}^{\bullet}$$

$$H_{k,l}^s = G_{1,l} \frac{X_{1,l+1}^s}{X_{1,l+1}^{\bullet}}$$

$$X_{k,l}^s \rightarrow X_{k,l}^s + H_{k,l}^s$$

$$X_{k,l+1}^s \rightarrow X_{k,l+1}^s + H_{k,l}^s$$

This adjustment is made first for  $l = 1$  and then for  $l = 2$ .  $G_{k,l}$  Since differential loss will always be greater from group 1 compared to 2 and 3, and group 2 compared to group 3,  $G_{k,l}$  is always positive.

*(ii) Calculating the force of infection*

The force of infection is calculated on the basis of the rate at which individuals change sexual partner, HIV prevalence among their sexual partners, the number of sex acts in each partnerships and the use of condoms.

Individuals in each sex and sexual activity group form partnerships at a set rate:  $c_{k,l}$ , which is parameterised by a mean rate of partner change for that gender ( $\bar{c}_k(t)$ ) and two parameters that give the relative partner change rate for those in the highest ( $\varpi_{k,1}$ ) and next-highest ( $\varpi_{k,2}$ ) risk groups relative to those in the lowest risk group (so that  $\varpi_{k,3} = 1$ ). The partner change rates for each group are thus calculated as:

$$c_{k,l} = \varpi_{k,l} \frac{\bar{c}_k}{\sum_l \phi_{k,l} \varpi_{k,l}}$$

$W_{k,l}$  is the total number of sexual partnership offered by individuals in that gender and sexual activity group:

$$W_{k,l} = c_{k,l} \sum_s X_{k,l}^s$$

Men and women form partnerships so that a fraction,  $\varepsilon$ , of their partnerships are directed only to those in those of the opposite gender in the corresponding sexual activity group. The rest are distributed randomly among those of the opposite gender, according to the number of partnerships available. For men and women,  $\rho_{k,l,l'}$  is the fraction of partnerships that individuals in the  $k^{\text{th}}$  gender and  $l^{\text{th}}$  activity-group form with individuals of the opposite gender in the  $l'^{\text{th}}$  activity group (the prime denotes that the index relates to those of the opposite gender).

$$\rho_{k,l,l'} = \varepsilon \delta_{l,l'} + (1 - \varepsilon) \frac{W_{k,l'}}{\sum_{l'} W_{k',l'}}$$

Here,  $\delta_{i,j}$  is the Kronecker delta:  $\delta_{i,j} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$

To ensure that the total number of partnerships formed by men and women are consistent, the following correction is made:

$$D_{a,b} = \frac{W_{2,l}c_{2,l}\rho_{2,l,l'}}{W_{1,l}c_{1,l}\rho_{1,l,l'}}$$

$$\rho_{2,l,l'} \rightarrow \rho_{2,l,l'}(D_{a,b})^{\theta-1}$$

$$\rho_{1,l,l'} \rightarrow \rho_{1,l,l'}(D_{a,b})^{\theta}$$

In this way,  $\theta$  determines the extent to which the pattern of partnership formation is determined by the parameters estimated from men's reported sexual behaviour.

In partnerships between individuals in sexual activity groups  $l$  and  $l'$ , the number of sex acts and the level of condom use is determined by whether the partnership is classified as 'regular' or 'casual'. If the partnership is formed among individuals from the two riskiest groups, it is classified as 'casual', otherwise it is classified 'regular'. Let  $L$  denote whether a type of partnership is casual or regular, so that:

$$L = \begin{cases} 1, & \max(l, l') < 3 \\ 2, & \text{otherwise} \end{cases}$$

The number of sex acts in casual and regular partnerships is, respectively,  $n_1$  and  $n_2$ ; the fraction of sex acts in which condom are used in casual and regular partnerships is, respectively,  $q_1$  and  $q_2$

The force of infection is calculated as:

$$\lambda_{k,l} = c_{k,l} \sum_{l'} \left( \rho_{k,l,l'} \left( 1 - (1 - p_{k',l'})^{(1-q_L)n_L} \right) \right)$$

Here,  $p_{k',l'}$  is the average chance of transmission per sex-act in sexual partnerships formed with individuals of that sex and sexual activity group. It depends on the prevalence and the stage of infection of those in that group in the following way:

$$p_{k',l'} = \frac{\sum_{s=1}^9 \beta_{k',s} X_{k',l'}^s}{X_{k',l'}^{\bullet}}$$

The chance of transmission per sex acts from individuals of sex  $k$  in disease state  $s$  ( $\beta_{k,s}$ ) is specified with three parameters: the average rate of transmission from individuals with latent infection, averaged across men and women ( $\beta$ ); the relative rates of transmission for each stage of disease relative to latent infection ( $\kappa_s$ ); and, the relative chance transmission from females versus transmission from males ( $\pi_1: \pi_2 = 1$ , by definition). The values for  $\kappa_s$  are estimated using observational data from rural Rakai. However, it possible that the assumed rate of transmission from individuals with pre-AIDS and AIDS has a large influence of the size of declines in HIV prevalence [97]. Since there remains uncertainty in the extent to which the rate of transmission increases as symptoms develop and because this may be associated with decreases in coital frequency or the rate of forming new partnerships, an additional parameter is introduced,  $h$ , which allows the relative rate of transmission from pre-AIDS and AIDS individuals to be adjusted independently of rate of the transmission at other stages.

$$\beta_{k,s} = \pi_k \kappa_s \left( \beta \frac{2}{1 + \pi_k} \right) \quad s = 1,2,3,4,5,9$$

$$\beta_{k,s} = h \pi_k \kappa_s \left( \beta \frac{2}{1 + \pi_k} \right) \quad s = 6,7,8$$

*(iii) Changes in sexual behaviour parameters*

Three sexual behaviour parameters are allowed to change in a piece-wise linear fashion; the mean rate of partner change for men, the mean rate of partner change for women, and the fraction of sex acts in which condoms are used in casual partnerships. These parameters all change at the same time, and this is parameterised by: (i) the time at which the changes in behaviour start,  $\zeta$ ; and, (ii) the time it takes for the behavioural parameter to reach its new value,  $F$ . The eventual relative change in the mean rate of partner change for women and men are denoted, respectively:  $\Delta_1$  and  $\Delta_2$ . The relative change in condom use in casual partnerships is denoted:  $\Delta_3$ . Thus:

Time ( $t$ )	Mean partner change rate: women and men	Condom use in casual partnerships
$t < \zeta_1$	$\bar{c}_k(t) = \bar{c}_k^*$	$q_2(t) = q_2^*$



$\zeta_1 \leq t < (\zeta_1 + F)$	$\bar{c}_k(t) = \bar{c}_k^* \left( 1 + \frac{(t - \zeta_1)}{F} (\Delta_k - 1) \right)$	$q_2(t) = q_2^* \left( 1 + \frac{(t - \zeta_1)}{F} (\Delta_3 - 1) \right)$
$t \geq (\zeta_1 + F)$	$\bar{c}_k(t) = \Delta_k \bar{c}_k^*$	$q_2(t) = \Delta_3 q_2^*$

Where a \* indicates the initial value for that parameter.

(iv) Model outputs

The key output from the model for comparison to data is the prevalence time-series. This is calculated by defining the calendar year when the simulation is assumed to start ( $T_0$ ) and recording prevalence, in the way shown below, at the mid-point of each subsequent year.

$$p(t + T_0) = \frac{\sum_{s=2}^9 X_{\cdot, \cdot}^s}{X_{\cdot, \cdot}}$$

However, although the model outputs simulated prevalence in the general population, it must be compared with data collected in ante-natal clinics (ANC). It has been shown that there are usually differences between prevalence measured in the general population and at the ANC [61], so the model output is *calibrated* to represent ANC prevalence.

$$\rho(t) = \Phi\left(\Phi^{-1}(p(t)) + d\right)$$

Where  $\Phi$  is the cumulative distribution function of the standard normal distribution and  $d$  is the 'calibration constant' parameter that is held constant within one simulation but can be varied between simulations (see next section).

Numbers of new infections in the period 1995-2008 are also calculated. The number of new infections in the simulation model is rescaled so that the denominator population matches the estimated population size and projected growth of the country of interest to provide consistency between derived estimates and those from other sources. Preliminary investigations showed that the size and the growth rate of the simulated population do not substantially affect prevalence or incidence rates. The population size in year 1995 is  $Pop_{1995}$ , and the growth rate (assumed constant) in the period 1995-2008 is  $\nu$ . The number of new infections is then calculated as:

$$I(t + T_0) = \frac{(Pop_{1995}) \exp(\nu * (t + T_0 - 1995))}{X_{\cdot, \cdot}} \sum_l \sum_k \lambda_{k,l}(t) X_{k,l}^1(t)$$

*(vi) Fixed Parameter Values*

*Values of Parameters That Do Not Vary Between Simulations*

Relative rates of transmission in the different stages and the time spent with acute infection are based on observational data from rural Uganda [59]. Durations of pre-AIDS and AIDS are based on a meta-analysis of survival rates in resource-poor settings [98]. In the model, individuals entering the pre-AIDS state are assumed to have a CD4 cell count that has decreased below 200 and are developing symptoms of immune-suppression. A meta-analysis [98] indicates that survival of infected individuals from the time when their CD4 cell-count is equal to 200 is approximately three-times as long as for patients with a CD4 count less than 200. Studies in resource-poor settings show that median survival for patients with CD4 counts below 200 is approx 1 year; therefore, we assume that the same relationship holds for resource-poor settings and estimate median survival for those with CD4 cell counts equal to 200 as 3 years. This three year period includes time spent in pre-AIDS and AIDS states. The same meta-analysis also indicates that mean survival time from developing AIDS to death is approximately 1 year; we therefore assume that average time spent with pre-AIDS conditions is 2 years. Duration of latent infection is estimated through fitting to reported rates of net HIV-associated mortality in resource-poor setting [96], subject to the constraint of durations of pre-AIDS and AIDS already discussed (Figure S1). The duration with pre-AIDS for those individuals that will receive ART is based on another modelling study focussed on the delivery of treatment relative to disease progression in Africa [99]. Survival on ART is based on extrapolating observed first-year mortality rates for individuals in resource-poor setting starting treatment [41,100].

Parameter		Value	Reference
Ratio of HIV transmission male-to-female versus female-to-male	$\pi_1$	2.00	[101,102]
Relative chance of HIV transmission (versus with latent infection): acute infection	$\kappa_1$	11.4	[59]
Relative chance of HIV transmission (versus with latent infection): pre-AIDS	$\kappa_6 = \kappa_8$	3.29	[59]
Relative chance of HIV transmission (versus with latent infection): AIDS	$\kappa_7$	6.14	[59]
Relative chance of HIV transmission (versus with latent infection): on treatment	$\kappa_9$	0.14	(23, 24: Assuming 50-100 viral copies per mm <sup>3</sup> )
Mean time with acute infection	$1/\sigma_2$	5 months	[59]
Mean time with pre-AIDS, and will not receive treatment	$1/\sigma_6$	2 years	[98]
Mean time with pre-AIDS, and will receive treatment	$1/\sigma_8$	2 months	[99]
Mean time with AIDS	$1/\sigma_7$	1 year	[98]
Mean time on treatment	$1/\sigma_9$	10 years	[100]
Year when ART first becomes available,	$a_{start}$	2003	[103]

**Table S1: Model parameters that have fixed values.**

*Prior Distribution for Parameters That Vary Between Simulations*

The location of priors for aspects of sexual behaviour relating to the individuals (mean rates of partner change, condom use in casual partnerships), is informed by data from the surveys, where specific information is not available, from other cohort studies, not necessarily local to that target setting [29,104]. The ranges of prior distribution is expected to extend beyond the sampling uncertainty in these surveys, so credible limits for

some sexual behaviour parameter values may be defined by examining the variance between population sub-groups within the survey.

Priors for parameters relating to the higher order properties of the sexual partner network (sizes and relative rates of partner change in the sexual activity groups, balancing, mixing and replacement) can be agnostic, that is uniform distribution over the entire range of feasible values.

The location of prior for the rate of transmission during latent infection reflects observational data from rural Uganda [59], but gives greater support to higher values than indicated in that study. This is done in recognition of two factors: first, the observational data likely underestimate the transmission probability per sex act due to inaccurate reporting of number of sex acts and through the selection bias of only including sero-discordant couples; second, it is known that sexually transmitted disease can increase the chance of HIV acquisition and transmission [105], but since they are not explicitly included in the model, higher value for the transmission probability of HIV can implicitly capture their basic effect.

**Supplementary Appendix 2. Payout profiles of payout functions from Table 1.**

The first two panels in this supplementary appendix repeat for reference Figure 11 of the text. The remaining panels display the payout profiles for the other nine payout functions of Table 1. Each payout function is presented for two sample sizes,  $N=10,000$  and  $N=40,000$ .

As described in the text, payout functions 1 through 6 are threshold functions, while functions 7 through 10 are continuous functions.

