



**HIV** Modelling  
Consortium

# **Strengthening The Use of Mathematical Models in Community Trials**

**Wednesday 17 – Thursday 18 October 2012**

**Boston, Massachusetts, USA**

**Meeting Report**

# Strengthening The Use of Mathematical Models in Community Trials

Wednesday 17 – Thursday 18 October 2012

Boston, MA, USA

## Meeting Aims and Context

### Aims

1. Extensively review models for community trial design, share and discuss plans for future model development and form plans for future collaborations (cross-country/cross-trial comparison).
2. Develop an analysis protocol that will enable a rigorous review of initial model predictions to be compared with eventual trial results.
3. Create a closer alignment between statisticians and modellers to enable best conduct of community trials and develop the field to allow integration of modeling into future trials.

### Context

In September 2011, PEPFAR announced awards for a new initiative totalling \$45 million over four years to examine the effectiveness of combination approaches to HIV prevention. Three studies received support from this initiative:

- **The PopART trial**, funded by the NIH, and led by the London School of Hygiene and Tropical Medicine (LSHTM) will examine a strategy linking household-based HIV testing to universal community-based HIV treatment in Zambia and South Africa.
- **The Botswana trial**, funded by the CDC and led by the Harvard School of Public Health (HSPH) will evaluate the impact on HIV incidence of expanding population coverage of an integrated set of HIV prevention interventions.
- **The Iringa trial**, funded through an existing USAID award and led by Johns Hopkins University (JHU) will evaluate the impact of an integrated set of biomedical, behavioural and structural prevention interventions to reduce HIV incidence in the Iringa region of Tanzania.

The design of these three trials is described in Table 1. For the first time, an HIV transmission dynamic modelling component has been incorporated in these trials to complement statistical analyses needed to inform trial design, interim evaluation and interpretation of the final results in terms of short and long term impact.

This meeting was organised by the HIV Modelling Consortium to review the models and to discuss strategies for their evaluation to strengthen trials and vice versa. This is a unique opportunity for progress in the field of HIV modelling and early discussions are needed to ensure the benefits are maximized.

Study	CDC/HSPH <sup>a</sup>	JHU/USAID	PopART (HPTN 071)
<b>Site</b>	Botswana	Iringa, Tanzania	Zambia+South Africa (Western Cape)
<b>Number of arms</b>	2	2	3
<b>Intervention arm(s)</b>	A: Enhanced HIV testing (including mobile and home-based testing), active linkage to care and treatment; enhanced MC; ART for all HIV-infected persons with CD4<350 cells/μl or with HIV-1 RNA>10,000 copies/ml; and point-of-care CD4 testing in antenatal clinics with universal HAART in pregnancy started by 28 wk gestation, as well as HIV retesting at delivery among women HIV-negative in second trimester or earlier	A: Treatment by CD4<350 cells/μl; active scale-up and linkage to MC; cash transfer for young women; targeted outreach to the most at-risk populations (including female sex workers); social and behaviour change communication	A: Universal community home-based testing; active linkage of HIV-positive individuals to care and immediate ART according to national guidelines and/or MC. B: Same as A but ART at CD4<350 cells/μl
<b>Control arm</b>	B: Standard of care <sup>b</sup>	B: Standard of care <sup>c</sup>	C: Enhanced standard of care <sup>d</sup>
<b>Design</b>	Pair matched	Stratified	Triplet matched
<b>Number of randomized clusters</b>			
Total	30	24	24 (South Africa: 9, Zambia: 15)
Per arm	15	12	8
<b>Average size of randomized cluster</b>	5,800	8,000–10,000 (~55%>15 y)	50,000 (25,000>18 y)
<b>Overall cohort followed up</b>			
Age eligibility	16–64 y	15–39 y	18–44 y
Size per cluster	~500 adults per cluster	~500 adults per cluster	~2,500 adults per cluster
Total size	15,000	12,000	60,000
<b>Primary Outcome</b>	HIV incidence <sup>e</sup>	HIV incidence <sup>e</sup>	HIV incidence <sup>e</sup>
<b>Follow-up duration</b>	3–4 y	2 y	2 y
<b>HIV incidence assumption</b>	~1.5 per 100 person-years	1.0–1.5 per 100 person-years	1.0–1.5 per 100 person-years
<b>Anticipated HIV prevalence at baseline</b>	25%	10%–15%	15%
<b>Target reduction in incidence</b>	In arm A versus B: ~50%	In arm A versus B: ~40% (35%–50%)	In arm A versus C: –50% to 60%; in arm B versus C: –25% to 30%
<b>Stages when modelling is currently planned</b>	Start	Start, interim, final	Start, final
<b>Status</b>	Planning	Pre-trial	Pre-trial

Data as of 15 March 2012.

<sup>a</sup>The design of the intervention and plan of analysis for this trial are still being finalised.

<sup>b</sup>Standard of care is ART for HIV-positive individuals with CD4<350 cells/μl or AIDS.

<sup>c</sup>Standard of care is standard referral to MC and ART according to Tanzania guidelines (this will soon change from CD4<200 cells/μl to CD4<350 cells/μl, initially focusing on HIV-positive people with tuberculosis and pregnant women).

<sup>d</sup>Standard of care is no home-based testing or home-based visit to facilitate linkage to ART. ART given according to country guidelines; standard referral to MC.

<sup>e</sup>Cumulative HIV incidence measured over the trial duration.

CDC/HSPH, US Centers for Disease Control and Prevention/Harvard School of Public Health; HAART, highly active ART; JHU/USAID, Johns Hopkins University/United States Agency of International Development; PopART (HPTN 071), HIV Prevention Trials Network.

doi:10.1371/journal.pmed.1001250.t001

**Table 1. Main characteristics of cluster randomized controlled trials for combination prevention of HIV transmission commissioned by PEPFAR (1).**

## The HIV Modelling Consortium

The HIV Modelling Consortium, funded by the Bill & Melinda Gates Foundation via a grant to Imperial College London, aims to help improve scientific support for decision making by co-coordinating a wide range of research activities in mathematical modelling of the HIV epidemic.

This involves:

- Identifying questions that demand mathematical modelling input and identifying new modelling results that may require further validation.
- Facilitating sharing of information, modelling techniques, data and expertise between research groups.
- Providing a forum for rigorous review of new mathematical modelling research and tools.
- Providing funding through sub-contracts to commission research to address those needs.
- A steering committee of leaders in HIV programme and policy directs the focus of the work of the consortium. Further information on the HIV Modelling Consortium is available in a standard briefing document and information about other work packages undertaken by the HIV Modelling Consortium is available at the website [www.hivmodelling.org](http://www.hivmodelling.org)

## **Session 1: Review of trial models and plans for future models.**

### **PopART model (by Anne Cori and Christophe Fraser at ICL)**

The PopART model is a deterministic compartmental model representing individuals aged 18 to 55 structured by sex, circumcision status, risk group, HIV status, clinical stage and treatment status. Risk groups are defined based on their partner change rates and mixing between these is fitted to data. Five clinical stages are included: acute infection and CD4 count >500, between 500-350, between 350-200 and <200 cells/mm<sup>3</sup>. Treatment is assumed to be scaled up from 2004 at CD4<200 and the rate of treatment for those with CD4 count between 200-350 is assumed to be half of that for those with CD4 count<200. There is no treatment for those with CD4>350. Transmission efficacy is modelled as dependent on CD4 cell count.

Contamination from other clusters is represented by assuming 5% of contacts occur with neighbouring communities. In order to identify parameters that have the strongest input on results, sensitivity analyses will be carried out by varying parameters associated with the scale up and effectiveness of the different components of the intervention. At the moment delays in the treatment cascade are not included and this is the priority in terms of changes to be made to the model.

The delay to reach undetectable viral loads is not included in the model at this stage and this will result in a delay in the outcome (incidence reduction). There was concern about how this could affect the power of the trial. It is thought that this should not affect the power in year 2 but it might decrease power in year 1.

It was also noted that progression from CD4 cell count >500 to CD4 cell count between 500 and 350 was assumed to be faster for those receiving treatment than for those not on treatment. This is an error that will be fixed in the next version of the model. When asked whether the model would be fitted to each site, it was said this was part of the plan but that differences between sites would be treated as a question rather than as a baseline assumption (part of the issues that may explain differences in the effectiveness of the intervention).

There was discussion around which assumptions would have the strongest impact on the short-term results: those determining distribution by clinical stages and their infectiousness or those determining sexual behaviour. The model assumes that limited behavior change will occur from counselling but conversely is quite optimistic about the effect of treatment since delays in the treatment cascade are not included. These are issues that will have to be taken into consideration for the analyses to come.

There was discussion around the possibility of including phylogenetic studies to determine the effect of contamination but this would cause problems with the IRB as at the moment the HIV test results are unlinked.

### **Iringa Trial model (by Mike Pickles *at al.* at ICL and LSHTM)**

Iringa is an area of extremely high HIV prevalence in comparison to Tanzania as a nation. The trial has been designed for 2 years with a possibility of further extension. The modeling component of the trial differs to PopART as it was incorporated once the trial was designed. Two models were built to answer different questions about the trial:

#### Model 1:

The aim of this model is to estimate the coverage and length of the intervention needed to obtain the expected reduction in incidence.

A simple deterministic model stratified by gender that represents the overall sexually active population aged 15-49. Only heterosexual relationships are considered. Susceptible women are divided in three groups: 15-24, 25-49 and female sex workers (FSW). Susceptible men are divided in three groups: circumcised, not circumcised, healing following operation. Infectivity and life expectancy vary by stages of HIV infection and treatment status. A delay is considered for eligible HIV patients to receive treatment. Currently eligibility is set at CD4 count <350 and <200 for the intervention and control arm respectively. Mortality on ART and drop-out are independent of time spent on ART.

Regarding sexual mixing some men can have sex with FSWs and contamination is considered by assuming some women have sex with men outside the community. The model was validated using a different model that originally simulated the impact of treatment in Zambia and South Africa and the results were similar.

The results suggested that for the trial to be successful in two years at least 94% of the population would need to be tested for HIV once (64% in first 6 months) and 80% of eligible patients should receive ART. In a four year trial 55% would need to be tested in year 1. This strongly argues for a trial extension of at least 1 year.

#### Model 2:

The aim of this model is to estimate the effect in changes in ART treatment criteria by answering the following questions:

- What happens if we change the CD4 eligibility criterion in the intervention arm?
- What happens when the eligibility changes in the comparison arm?

Incidence was obtained for 2 different eligibility criteria for the intervention arm (CD4<350 and <500) and the control arm (CD4<200 and <350). The results are consistent with the previous model suggesting it would take approximately 3 years to see a drop in incidence with CD4 eligibility <350 when about 70% are tested for HIV. Increasing the eligibility to 500 CD4 cells reduces incidence by ~10% and would minimize the required trial duration by nearly a year.

In conclusion, both models suggest the current trial duration of 2 years is unlikely to show an impact even with optimistic assumptions. A three-year trial may be the minimum and would require a fast scale-up, a well-functioning treatment cascade and high adherence. Reaching FSWs and other key populations remains important. Contamination can be a big problem but it is hard to quantify it. Data on

current ART coverage in Iringa is challenging. Clinic data are likely to be substantially over-estimated due to dropout and people registering at more than one clinic.

The model includes 2 pathways to initiate treatment:

1. Individuals are diagnosed with HIV and monitored until they are eligible for treatment according to the treatment guidelines in place
2. Individuals will access the health services later when they present with advanced and often symptomatic infection.

The proportion initiating treatment through each of these pathways will have an impact on the model results as it defines the number of people receiving early treatment (pathway 1).

Contamination between clusters could be an issue within this trial as Iringa is a central trucking area with a high number of FSW. This effect may result in dilution and is therefore must be considered when reviewing model outputs. Actions to remedy this are underway. First, extensive field interviews with truck drivers and clients are being conducted to discuss their sexual partners in detail. Second, phylogenetic samples are being taken but this is expensive and can only inform on partnerships of HIV positive people. No gold standard exists for measuring contamination and is therefore an important issue to be addressed.

#### **CDC / Botswana trial model (by Riu Wang et al. at HSPH)**

The Botswana trial utilizes a “cumulative network model” which first generates sexual networks and then simulates disease spread and intervention impact. At present no correlation between the number of partners and the duration of relationships is assumed. Each individual is defined by sex, transmission risk, disease progression, condom use, linkage to care and circumcision status. Mixing between communities is simulated by changing the proportion of contacts that are made outside the community, at present this is estimated to be at 20%. The simulation starts at the beginning of the trial and so the initial conditions for HIV prevalence, condom use, proportion of males circumcised, viral load and CD4 cell distribution and ART coverage correspond to what was observed at that time. The control arm is optimistic (assumes rapid scale up of circumcision and ART coverage at CD4<350) as it is designed to simulate the health system if all planned improvements are incurred. Sensitivity analyses were conducted to investigate the effect of lower coverage or slower scale up of the. The low incidence in the standard of care community was also taken into account in the analysis as this might lead to futility of the intervention.

As this is a standing start model (i.e. the simulation begins at the same time as the event it is modeling, in this case the trial, and does not include the history of progression of the epidemic) there are concerns about model validation. A recommended method of validation would be to utilize the model’s ability to replicate past prevalence and incidence trends however historical data might not be informative. An alternative method would be to model the 4 years prior to trial initiation for which there might be better data. Concerns were also raised about the possibility that the model has not reached equilibrium, which

could have an impact on the results: if incidence plummets after 3 years when reproducing historical data this would suggest the estimated effect of the intervention could simply be driven by the model mechanics. However, the Botswana model has been tested against the Goals model (Futures Institute) which does include historic data about progression of the epidemic and it matched the CD4 <500 data for all 4 years of the study. Additionally, since the purpose of the exercise was to model intervention impact over a short time period, this should not be a problem. Conversely, it was argued that in this case a simpler model might have been more appropriate and easier to interpret.

The purpose of the Botswana model may differ to that of PopART and Iringa as it aims to determine whether it would be “plausible” to see an effect while the other two aim to explain why we could observe an effect or not. For trial impact estimation having a dynamic model might not be absolutely necessary because of the short time scale. Checking the impact of including non-linearities would help answer this question. The modellers from Iringa had done an informal check by multiplying effects and found a 20% difference approximately. The Botswana trial model could be used to identify the parameters that are driving the results. It might be that less information is required than previously thought.

It was noted that a greater marginal return was obtained for this trial than the PopART and Iringa trials. This large effect seems to be mainly attributed to treating people with VL>10,000 copies/mL which was surprising to some. Sensitivity analyses are needed to determine what is driving the decrease in incidence. The viral load treatment initiation threshold was reduced from 50,000 copies/mL to 10,000 copies/mL allowing more individuals to enter the eligibility criteria. This ensures individuals with high viral loads who are more infectious and therefore more likely to transmit are entered into care. Uncertainty exists around what the minimum threshold for transmission is which would impact the model results depending on the assumptions. Additionally higher viral load is associated with a faster decline in CD4 count but this is not included in the model possibly resulting in an overestimation of the intervention impact.

There were also discussions around the data used to parameterize the model. Data to inform the duration of relationships comes from the Mochudi partnership survey where it was suspected that people tended to underreport number of partners. The Likoma island study was much more thorough, and participants were asked to name partners and provide follow up details. Often people are much more reluctant to report relationships when they are longer term. Assessing the severity of the bias and determining how to correct for it is important. Getting the two data sets together and comparing them would provide an initial impression of the potential biases. In the model complete independence between the number of partnerships and the duration of these was assumed. Investigating how this affects the estimate of the intervention impact was considered a priority.

Data on mixing patterns is sparse and this could be a determining factor for the model’s estimates as well as the success of the intervention. The Africa Centre has investigated partnership formation in relation to geography. There appears to be a maximum threshold of 10km for partner formation with



most individuals mixing within the vicinity. It remains unclear if geography can be a proxy for risk behaviours. The investigators expressed reservations regarding the validity of the data at this stage.

### **General discussion**

A few issues were common to all models:

- The importance of modelling delays and leaks in the cascade is a priority as it will influence estimates of the effect size. These are not as important for the long-term impact but are critical to estimating the effects occurring in the months after the intervention.
- All models have considered “contamination” but it might be more influential than previously thought, especially in Iringa due to the mobile nature of its population. Efforts to obtain empirical data on this are needed.

## **Session 2: Preparation of model protocols for archiving (for future model comparison)**

Mathematical models are considered useful tools for anticipating the impact of HIV interventions. However, very few model projections have been validated with data and so there is little evidence supporting this statement. The inclusion of a modelling component at such an early stage in these 3 trials (4 when including the Africa Centre trial) provides an opportunity to formally establish a validation process.

The model estimates of reduction in incidence are recorded in the trial design protocols (at least in those that included modelling at the trial design stage) and it will be possible to compare them to data. The model estimates could be far from the actual results due to differences between the expected and achieved programmatic variables. For instance, if only half of the planned coverage was attained, incidence reductions are likely to be smaller than originally expected. The model estimates could also differ from the data because of uncertainties in parameter values relating to either the natural history of infection, sexual behaviour or population demography or in the structural assumptions. Identifying these sources of error will allow improvement of future models and, possibly, differentiation between accurate and inaccurate approaches. It is essential to know whether models helped during the trial and if they did to show when and why they provided good guidance. Keeping track of why the models did not work is also part of the scientific process. Models are used by governments, institutions and now RCTs and they are becoming increasingly influential at the decision making level. It was agreed that it is essential to promote progress and innovation in this field, to make sure the models that are developed in 4 years are better than today's.

## Process

In order to ensure this exercise is taken to its full potential it is essential to plan it formally and decide on some fundamental questions ahead of the trial results.

The key to the process is to formally define the “models” that were used at the beginning of the trials (or could have been used at the beginning of the trials). The model protocols (that define all aspects of the model, parameters and analysis) will be documented and archived in a secure online site that is managed by a third party.

It is also necessary to define strict boundaries in the definition of the model between the “structure”, the “parameters” (epidemiological and biological assumptions) and the “intervention assumptions” (a link to the full protocol can be found below).

Each group presented their idea on the process to be followed and a few themes emerged from the discussion:

### ***When is the appropriate time to “freeze”?***

Both additional time and information will allow making the model more accurate and so at the end of the trial this will be different from the model used initially to inform the power calculations. When to freeze the model for comparison to the trial results is one of the questions that emerged during the discussion. It was agreed that this decision depended on the purpose of the model, which might also change along the different stages of the trial. The models are initially used to inform the power calculations and give guidance on whether the aimed coverage and length of the trial will be sufficient to show an effect. Later on they can be updated with data obtained from the field to evaluate the observed reductions in incidence in the light of the programmatic indicators. At the end of the trial, models can be used to explain the results, whether these are positive or not. Based on this it was proposed that the model could be frozen at different stages during the process and validated against the questions asked at each of these stages. This approach would also allow learning about when models are useful in the context of trials. It might be that when data are scarce (i.e. at the beginning of trials), models are not informative for instance. It will be essential to be clear on the questions that are answered by each of these models to make sure the tests are well defined before the validation takes place.

### ***What aspect of the model constitutes the frozen part and what can we update?***

Initial model estimates are based on desired programmatic achievements. Therefore, if these programmatic aims are not reached, the model estimates are likely to be incorrect irrespective of how accurately the model reproduces the epidemic in the population of interest. Considering this, it is important to allow certain parameters to be updated with values obtained from the field. However, determining the parameters that fall in this category is a subjective issue. While some such as coverage of the intervention appear obvious, others such as adherence or behaviour change are more debatable. To decide what models will be made accountable for, it will again be important to specify what question

is being used as the test. Modellers will be asked to propose clear boundaries between types of assumptions in the models and for these distinctions to be made a part of the “frozen” protocol.

### ***Practicalities***

Questions around the practical implementation of this project were approached. Deciding what to document and how to store it are two fundamental issues. The PopART modellers have been using an interface (SourceForge) to record all changes made to the model through time. By keeping track of all changes, the entire process is documented which serves two main purposes: first it ensures transparency, second it provides a platform to study model development in the context of trials and learn from it.

Both of these aims, transparency and learning, are achieved if others can access and use the models. The PopART model is accessible to all and a few selected users can make changes to the code. The interface allows having several branches within which the code changes independently. This is useful to answer specific questions that need changes in the structure that would not be relevant to the main questions. Additionally, as it is possible that the researchers developing the tools today will not be working on this project in four years, the model code should be clearly annotated and functional so that the task of investigating the results can be done by others.

It was agreed that the models would be open access and that the Modelling Consortium website would be used as an interface to implement this. This will also allow to announce the project publicly and to formalize the exercise. However, some expressed concerns about other groups publishing ahead of them using their model and data from the trial as soon as it becomes available. It was decided to plan a time-limit and apportion credit.

To view the Freezer Protocol please visit: <http://www.hivmodelling.org/announcements/call-submissions-model-validation-exercise-community-trials>

## **Session 3: Modellers respond to questions coming up from working groups**

Within the PEPFAR Combination Prevention Coordination activities there are four workgroups: Laboratory, Questionnaire, Quantitative and Economic Evaluation. Several groups will be investigating the impact of the intervention beyond the study period and may require input from the modelling groups. In particular the economic evaluation group will need estimates of the effectiveness of the intervention. Several questions arise when considering this eventuality

- How far out should model projections go (in terms of time and in generations of infections)?
- How much will varying counterfactuals impact the comparability of the impact estimates?
- Will the models estimate both the treatment and prevention benefits?
- How will the effects of the different components (especially in light of the enhanced linkages between components) be modeled (this important given that the studies are not powered to empirically answer contribution of different components)?
- How the epidemiological models will take into account quality of life?

There were discussions as to how far ahead model projections should go. The effect of ART on transmission can be observed on a relatively short time scale but its effect on survival may require much longer to be observed. The same applies to other indicators such as ART treatment-years averted, incidence of Tuberculosis or proportion of children that become orphans as a result of AIDS. Timescales such as 10 years and 25 years might be required to measure intervention impact on some of these indicators. However, modellers agreed that the longer the time scale, the greater the number of assumptions and therefore the greater the uncertainty in model outputs. Demographic processes as well as changes in behaviour are likely to modify the results and so caution must be taken when producing analyses about intervention impact in the long term. Conversely it was raised that the perspective of policy makers also had to be taken into account and that these rarely make decisions on a 25-year basis. It would be more pertinent to consider other questions that might be more relevant to policy makers such as the generalizability of trial results. Models might be able to translate these to other epidemic scenarios so that the trial findings can be rapidly applied elsewhere. Data that helps answering this question would be a very valuable resource.

Establishing a dialogue with the other groups and obtaining a list of required model outputs will be essential to evaluate feasibility and to determine the data to be collected to produce those outputs. It was suggested that the HIVMC could be called upon to act as the facilitator between the different groups and ensure the participation of modellers to the relevant calls and meetings.

#### **Session 4: Discussion to identify opportunities for best involvement of modelling in the trials**

How the models could be used to inform the trial at a later stage is still to be clearly determined. A scenario in which models might be most useful is a failure to observe the desired effect size after a certain period, which might lead to stopping the trial for futility. Interim analyses could evaluate if an improvement of the intervention performance would result in an effect by the end of the trial and if so, provide clear programmatic goals. It could also bring evidence for the need to prolong the trial in case the analyses showed a high likelihood of observing an effect later on.

The capacity of the models to guide decision-making during the trial will also depend on the data that will be made accessible to modellers from the trial and on the timing of this. Due to confidentiality issues this might be quite restricted and modellers will have to work along the Data, Safety and Monitoring Board (DSMB) schedule.

Data on sexual behaviour would be useful for these interim analyses. Questions that are relevant to the modellers perspective could be included in the participants' yearly clinic visit. Choosing the issues to be investigated should be based on sensitivity analyses and efforts should be focused on those for which

there is most uncertainty. Information on concurrency and assortativity in mixing and contamination are likely to be of particular importance. There will always be a problem around the accuracy of reported sexual behaviours and the first way to test this is by verifying that male and female data match.

When these interim analyses will be carried out is also a question. In the case of Iringa, there is a plan to carry the first interim analysis at 9 months and 2 years after the trial start. It cannot be too soon after because not much more information will be available and modellers also need time to carry out the analyses but it cannot be too late either as a decision to extend the trial must be made about a year ahead as the logistics of it are complex and involve several players. Both PopART and the Botswana trials are similar in that the trial is planned over three years but funding is only confirmed for the first two years. The decision for a third year of funding will involve a milestone review of the two first years, basically assessing whether the trials should be stopped on the grounds of futility. The modelling component in these scenarios might bring valuable information to guide the decision.

Finally, how modelling contributions will be integrated in the decision process is also to be determined. The DSMB does not traditionally include a modelling expert but this will be required to ensure that the additional information brought in by models is well evaluated and considered. The Africa Centre trial has already had its first DSMB meeting and has hardwired the mathematical modelling component. The other groups could learn from this experience.

## **Session 5: Debate on broad areas intersection between epidemiological modelling, statistics and trials**

**Question 1:** Are models that are used to make projections of the impact of Treatment on HIV incidence based on poorly backed-up priors that have a strong influence on results? If so, how should we use modelling results before trial data are available?

There was a general agreement on the fact that models are making incidence projections using poorly backed up priors. There is uncertainty in terms of sexual behaviour (sexual mixing/contamination, assortativity, partnership duration, behaviour change), natural history of infection (acute infection, MMC & infectiousness, heterogeneity in susceptibility, disease progression with and without treatment, drug resistance); demography (migration) and programmatic indicators (baseline ART coverage, potential scale up of ART in SoC arm, expected speed of scale up of intervention). However, how these four main sources of uncertainty influence the predicted impact of trials and the relative importance of each is unclear. Some work suggests that ART interventions are not too sensitive to sexual behaviour and natural history of infection assumptions (i.e. model comparison of ART effect in South Africa) especially when these are major interventions and we measure their short-term impact.

The best way to minimise this problem is by ensuring models are capable of reproducing past trends and by limiting the analyses to short time projections. Models will not provide an exact estimate of the

intervention effect but can provide a range of likely effect sizes as well as identify a likely time course of this effect. They can guide programmatic decisions (level and speed of scale-up/ intensity) as well as budget planning and alert about potential threats to the trial (contamination, slow scale-up, insufficient duration).

Sensitivity analyses must be used to identify important levers for the implementers to consider (speed of scale-up, retention, adherence) and design sub-studies to collect additional data.

**Question 2:** With the data from the trials, will we end up with models that, once fitted to the outcome of the trial, that will be less exposed to informative but poorly backed up priors? And if not, what can be done about it?

Some argue that data from the trials will provide information on some important programmatic parameters including retention in care and adherence and should also shed light on some key uncertainties, especially if phylogenetics and methods minimizing reporting bias are used (infections from outside clusters, sexual mixing, stage of infection, who-infects-whom). If model parameters cannot be measured directly the use of Bayesian inference should allow us to estimate them by the end of the trial in comparison to biological outcomes. However others felt that uncertainties would remain as the trials have limitations and do not address all questions. As per previous debates short-term impact prediction is not overly reliant on the difficult issues that are most contentious. Data and model triangulation will be useful in this respect. Increasing the number of outcomes should also allow reaching a higher degree of precision.

Finally the trials will help us make stronger predictions for the immediate impact of interventions similar to those found in the trials but the drivers of long-term impact are different to those for short-term so we will still be limited in our knowledge of the long term processes.

**Question 3:** In what ways can we ensure that the maximal value from these trials is extracted? In four years' time, what would be most disappointing outcome?

To make the most of these trials, the modelling analysis must be pre-planned and modellers must work in conjunction with the trial statisticians and triangulate their results with data.

Model results should be compared across trials – (especially if models do not pass validation in final analysis) and fitted to each other's data if required. Additional modelling analysis post trial would allow building on what has been learnt and improving methods of integrating model in trials

The most disappointing outcome would be futility that we cannot explain with the data available. Insufficient process/intermediate data to understand mechanisms of effect action could severely undermine the use of models to explain the results and to learn from them. It was felt it would be especially disappointing if trials were stopped prematurely when they would actually have achieved an effect and models were not able to predict this. In regards to the trials, the most disappointing outcome would be a merging of hypotheses across them due to changes in the national guidelines (loss of control arms or of different scenarios)

## Outcomes

### Aim 1: Model Review

Priorities were identified and trial modellers are in the process of developing the second version of the model which will be used in the interim analyses.

### Aim 2: Freezing the model

There was enthusiasm about the project and the idea of making the models open access. Clarity around a set of questions was achieved but the protocol still requires further development. A concept note will be produced by the HIVMC to set the basis and engage further discussion.

### Aim 3: Closer alignment between statisticians and modellers

It was agreed that efforts need to be made to ensure the maximum benefits are taken from past studies and especially from surveillance data. The importance of specific behavioural outcomes such as assortativity in mixing, concurrency and contamination were highlighted. Obtaining better data on these issues and improving estimates of uncertainty around them would enhance the quality of the models' projections. Alternative ways to investigate some of these issues such as through the use of geography and space seem promising and should be considered further. The notion of a data wish list for modellers that would be shared between the groups was proposed and will be promoted by the HIVMC. Finally, a modeller or a statistician with a modelling background is needed on the DSMB to make sure the contributions of modelling are well interpreted and considered.

### References:

1. Boily M-C, Mâsse B, Alsallaq R, Padian NS, Eaton JW, et al. (2012) HIV Treatment as Prevention: Considerations in the Design, Conduct, and Analysis of Cluster Randomized Controlled Trials of Combination HIV Prevention. PLoS Med 9(7): e1001250. doi:10.1371/journal.pmed.1001250

## Appendix

### 1. Agenda

OCTOBER 17 2012 – DAY 1	
09.00 -09.15	<b>Introductions</b> (Tim Hallett)
09.15 -12:00	<b>Review of trial models and plans for future models. Presentations from each trial model and sharing of full technical descriptions to provide open peer-review.</b> <i>Chair: Tim Hallett</i> PopART

	<ul style="list-style-type: none"> <li>- Christophe Fraser [POPART Project] – Hypothesis, design, status update (10 mins)</li> <li>- Anne Cori [POPART Project] – Model description (20 mins)</li> <li>- Discussion (20mins)</li> </ul> <p>Iringa Project</p> <ul style="list-style-type: none"> <li>- Noya Galai [Iringa Project] – Hypothesis, design, status update (10 mins)</li> <li>- Mike Pickes [Iringa Project] – Model description (20 mins)</li> <li>- Discussion (20 mins)</li> </ul> <p>CDC Botswana Project</p> <ul style="list-style-type: none"> <li>- Victor DeGruttola [CDC/Botswana Project] – Hypothesis, design, status update (10 mins)</li> <li>- Rui Wang [CDC/Botswana Project] – Hypothesis, design, status update (10 mins)</li> <li>- Discussion (20 mins)</li> </ul>
12:00 – 13:00	<b>Link up to LSHTM (Dial-in details provided on the day)</b>
13:00 - 14:00 Lunch.	
14:00 – 15:30	<p><b>Preparation of model protocols for archiving (for future model comparison)</b></p> <p><i>Chair: Josh Salomon</i></p> <ul style="list-style-type: none"> <li>- Tim Hallett – Mission and lessons learnt from other fields. (10 mins).</li> <li>- Anne Cori [POPART] (10 mins)</li> <li>- Mike Pickles [Iringa] (10 mins)</li> <li>- Rui Wang [CDC / Botswana Project] (10 mins)</li> <li>- General discussion on problems, pitfalls, plans and process. (30 mins)</li> </ul>
15.30 - 15.45: Coffee	
15:45 – 17:00	<p><b>Modellers respond to questions coming up from working groups</b></p> <p><i>Chair: Naline Sangrujee</i></p> <ul style="list-style-type: none"> <li>- How far out should model projections go?</li> <li>- How will the prevention benefit be separated from the treatment benefit?</li> <li>- How the epidemiological models will take into account quality of life? (How will QALYs be calculated?)</li> <li>- What counterfactuals we will use?</li> <li>- How we will model the effects of the different components (especially in light of the enhanced linkages between components)?</li> </ul>

OCTOBER 18 2012 – DAY 2	
09.00 - 12.00	<p><b>Model Validation Exercise Continued and Questions Arising</b></p> <p><i>Chair: Tim Hallett</i></p>
12:00 –	<b>Link up to LSHTM (Dial-in details provided on the day)</b>



13:00	
13:00 - 14:00 Lunch.	
14:00 – 15:00	<p><b>Discussion to identify opportunities for best involvement of modelling in the trials. Ideas have been raised</b></p> <p><i>Chair: Ben Masse</i></p> <ul style="list-style-type: none"> <li>- Speakers: Marie-Claude Boily, Victor De Gruttola, Christophe Fraser (10 mins each)</li> <li>- What questions are being raised as models are being integrated in trials? How will models be used as the trial is underway? Will models be used in no/no-go decision-making and how can this be strengthened? What lessons have already been learnt? Radical new suggestions? What research is required to move this forward?</li> </ul>
15.00 - 15.15: Coffee	
15:15 – 16:00	<p><b>Debate on broad areas intersection between epidemiological modelling, statistics and trials.</b></p> <p>Part I</p> <p><i>Discussion chair: Victor DeGruttola</i></p> <ul style="list-style-type: none"> <li>- Question: Are models that are used to make projections of the impact of Treatment on HIV incidence based on poorly backed-up priors that have a strong influence on results? If so, how should we use modelling results before trial data are available</li> </ul> <p>Part II</p> <p><i>Discussion chair: Christophe Fraser</i></p> <ul style="list-style-type: none"> <li>- Question: With the data from the trials, will we end up with models that, once fitted to the outcome of the trial, that will be less exposed to informative but poorly backed up priors? And if not, what can be done about it?</li> </ul> <p>Part II</p> <p><i>Discussion chair: Jim Hughes</i></p> <ul style="list-style-type: none"> <li>- Question: In what ways can we ensure that the maximal value from these trials is extracted? In four years time, what would be most disappointing outcome?</li> </ul>
16:00	<b>Meeting close</b>

## 2. Participant List

### PopART

**Christophe Fraser**

Imperial College London

**Richard Hayes**

London School of Hygiene and Tropical Medicine

**Anne Cori**

Department of Infectious Diseases, Imperial College  
London

**Sian Floyd**

London School of Hygiene and Tropical Medicine

### Iringa

**Marie-Claude Boily**

Department of Infectious Diseases, Imperial College  
London

**Noya Galai**

JHU

**Michael Pickles**

Department of Infectious Diseases, Imperial College  
London

### Harvard / Botswana

**Rui Wang**

Harvard School of Public Health, Harvard University

**Victoria DeGrutola**

Harvard School of Public Health, Harvard University

### Africa Centre

**Frank Tanser**

Africa Centre

**Ken Freedberg**

Harvard School of Public Health

### Additional Modelling and Stats Experts

**Josh Salomon**

Harvard School of Public Health, Harvard University

**Benoit Masse**

Statistical Center for HIV/AIDS Research &  
Prevention (SCHARP)

**Jan Hontelez**

Erasmus Medical Centre, Erasmus University

**Deborah Donnell**

SCHARP, Fred Hutchinson Cancer Research Centre

**Jim Hughes**

SCHARP, Fred Hutchinson Cancer Research Centre

**Johannes Berkhof**

University Medical Centre, Amsterdam

**Jamie Robbins**

Harvard University

**Dobromir Dimitrov**

SCHARP, Fred Hutchinson Cancer Research Centre

**Richard White**

TB consortium

**Till Barnighausen**

Africa Centre

### Funders / PEPFAR co-ordination team

**Geoff Garnett**

Bill and Melinda Gates Foundation

**Naline Sangrujee**

Centres for Disease Control

**David Burns**

National Institute of Health

### Secretariat

**Alex Welte**

SACEMA

**Tim Hallett**

Department of Infectious Diseases, Imperial College  
London



**Jeff Eaton**

Department of Infectious Diseases, Imperial College  
London

**Annick Borquez**

Mathematical Modelling in Community Trials Report

Department of Infectious Diseases, Imperial College  
London

**Kate Bilsborrow**

Department of Infectious Diseases, Imperial College  
London