

**HPTN 071-2 Phylogenetics in HPTN 071:
An ancillary study to “Population Effects of Antiretroviral Therapy to Reduce HIV Transmission
(PopART): A cluster-randomized trial of the impact of a combination prevention package on
population-level HIV incidence in Zambia and South Africa”**

A Non-IND Study of the HIV Prevention Trials Network

Sponsored by:

Division of AIDS, U.S. National Institute of Allergy and Infectious Diseases
U.S. National Institute of Mental Health
U.S. National Institutes of Health

Protocol Chair:

Christophe Fraser

Protocol Co-Chair:

Sarah Fidler

Version 1.0

15 January 2015

**HPTN 071-2 Phylogenetics in HPTN 071:
An ancillary study to “Population Effects of Antiretroviral Therapy to Reduce HIV Transmission
(PopART): A cluster-randomized trial of the impact of a combination prevention package on
population-level HIV incidence in Zambia and South Africa”**

TABLE OF CONTENTS

PROTOCOL TEAM ROSTER	4
SCHEMA	6
1. SUMMARY	8
2. BACKGROUND	8
3. STUDY OBJECTIVES AND METHODS	11
3.1 General approach.....	11
4. STAGES OF THE ANCILLARY STUDY	17
5. STUDY POPULATION	18
6. SAMPLING AND DATA COLLECTION	20
6.1. Sample collection.....	20
6.2. Target coverage and power calculations.....	23
6.3. Population cohort (PC) sampling	23
6.4. Health care facility (HCF) sampling.....	24
6.5. Estimated number of samples collected and identified for sequencing.....	25
7. LABORATORY PREPARATION AND TRAINING	29
8. SAMPLE ANALYSIS	30
9. ANALYSIS OF SEQUENCE DATA	31
10. ETHICS CONSIDERATIONS	32
10.1 Summary of ethical considerations.....	32
10.2 Community Engagement.....	33
10.3 Samples from PC participants.....	34
10.4 Samples from HCF patients.....	34
10.5 Community anonymization.....	34
11.0 SAFETY MONITORING AND SOCIAL HARM/ADVERSE EVENT REPORTING	35
11.1 Safety Reporting	35
12. ADMINISTRATIVE PROCEDURES	36
12.1 Protocol Registration.....	36
12.2 Study Activation	36
12.3 Study Coordination.....	36

12.4 Study Monitoring	37
12.5 Protocol Compliance.....	37
12.6 Study Discontinuation and Participant Termination	37
12.7 Investigator's Records.....	37
12.8 Use of Information and Publications	38
13. CONCLUSIONS.....	38
APPENDIX 1: SAMPLE INFORMED CONSENT FORMS.....	39
APPENDIX 1A: PC PARTICIPANT INFORMATION AND CONSENT FORM.....	40
APPENDIX 1B: HCF PARTICIPANT INFORMATION AND CONSENT FORM	45
APPENDIX 2: POWER CALCULATIONS	50
REFERENCES	53

PROTOCOL TEAM ROSTER

Helen Ayles

Clinical Scientist/Site Principal Investigator
ZAMBART
University of Zambia
School of Medicine, Ridgeway Campus
Lusaka, Zambia
Dept of Clinical Research
London School of Hygiene and Tropical
Medicine
Keppel Street
London WC1E 7HT, UK
Phone: +260 211 254710, 260 211257215
Email: helen@zambart.org.zm

David Burns

Chief, Clinical Prevention Research Branch
Division of AIDS, National Institute of
Allergy and Infectious Diseases
5601 Fishers Lane, Rm 4B40
Bethesda MD 20892, USA
Phone: +001-301-435-8896
Email: burnsda@niaid.nih.gov

Deborah Donnell

SDMC Biostatistician
SHARP-FHCRC
1100 Fairview Ave. North, LE-400
PO Box 19024
Seattle, WA 98109, USA
Phone: +001-206-667-5661
Email: deborah@ssharp.org

Bonnie Dye

Senior Clinical Research Manager
FHI360
359 Blackwell St. Suite 200
Durham NC 27701, USA
Phone: +1 919 544 7040 ext. 11379
Email: bdye@fhi360.org

Susan Eshleman

HPTN Network Laboratory
Johns Hopkins University School of
Medicine
720 Rutland Ave.
Ross Building, Room 646
Baltimore MD, 21205, USA
Phone: +001-410-614-4734
Email: seshlem@jhmi.edu

Sarah Fidler

Protocol Co-Chair
Imperial College London
CSL in Communicable Diseases
St Mary's Campus
HIV Clinical Trials Unit
Winston Churchill Wing
London W2 1NY, UK
Phone: +44 (0) 203 312 6790
Email: s.fidler@imperial.ac.uk

Christophe Fraser

HPTN 071-2 (Phylogenetics) Protocol Chair
Imperial College London
Department of Infectious Disease
Epidemiology
St. Mary's Campus
Norfolk Place
London W2 1PG, UK
Phone: +44 (0) 20 75943397
Email: c.fraser@imperial.ac.uk

Richard Hayes

HPTN 071 Protocol Chair
Department of Infectious Disease
Epidemiology
London School of Hygiene & Tropical
Medicine
Keppel Street
London WC1E 7HT, UK
Phone: +44 (0)20 7927 2243
Email: richard.hayes@lshtm.ac.uk

Paul Kellam

Virologist
Wellcome Trust Sanger Institute
Wellcome Trust Genome Campus
Hinxton
Cambridge CB10 1SA, UK
Phone: +44 (0)1223 494940
Email: pk5@sanger.ac.uk

Mohammad Limbada

Study Manager
ZAMBART Project
PO Box 50697
Ridgeway Campus
Lusaka, Zambia
ZAMBIA
Phone : +260-211-254710/+260-211-
257215
Email : mohammed@zambart.org.zm

Estelle Piwovar-Manning

HPTN Network Laboratory
600 North Wolfe St.
Pathology 306
Johns Hopkins University
Baltimore, MD 21287, USA
Phone: +001-410-614-6736
Fax: +001-410-614-0430
Email: epiwowa@jhmi.edu

Oliver Ratmann

Phylogenetics analyst
Imperial College London
Department of Infectious Disease
Epidemiology
St. Mary's Campus
Norfolk Place
London W2 1PG, UK
Phone: +44 (0) 20 75941451
Email: oliver.ratmann@imperial.ac.uk

SCHEMA

- Purpose:** To use phylogenetic methods to examine patterns of HIV transmission on a population level in selected communities in the HPTN 071 (PopART) trial and to identify factors associated with HIV transmission in these communities.
- Design:** The ancillary study will be performed in four stages:
- Stage 1: Sample and data collection.
 - Stage 2: Laboratory preparation and training.
 - Stage 3: Analysis of samples to obtain HIV sequence data.
 - Stage 4: Phylogenetic and statistical data analysis.
- Study Population:** Samples and data will be collected from two overlapping sources:
- Participants in the HPTN 071 Population Cohort (PC) in Zambia*
Samples will be collected from all PC participants as part of the main HPTN 071 trial. Nine communities in Zambia will participate in the phylogenetics ancillary study. At the 12-month (PC12) visit, all PC participants in the selected communities will be invited to participate in this ancillary study, irrespective of HIV status. No additional specimens will be collected from PC participants for the ancillary study; instead, a portion of the plasma stored in the main HPTN 071 will be used.
- Patients attending Health Care Facilities (HCFs) in the HPTN 071 study communities in Zambia*
Samples will be saved for this ancillary study that remain following CD4 T-cell count testing from patients undergoing routine CD4 cell count testing as part of their clinical management. This sampling will be performed at HCFs in the selected HPTN 071 study communities, for patients who agree to participate. Blood samples remaining after CD4 cell count testing, instead of being discarded, will be used to prepare a plasma aliquot for the phylogenetic study (i.e., no additional blood samples will be specifically collected for this study).
- Study Size:** Phylogenetic analyses performed in the ancillary study are expected to include sequence data from ~9200 individuals, including ~1200 PC participants and ~8000 HCF patients.
- Treatment Regimen:** This ancillary study does not include any treatment or intervention.
- Study Duration:** Data collection is expected to begin in 2015. Laboratory analysis of HCF samples is expected to begin soon after sample collection and continue until a year after sample collection ends, ~2017. Laboratory analyses of the PC samples is expected to begin after the laboratory assessments for PC in the main HPTN 071 study have been completed (~2019) and is expected to continue for ~2.5 years

Primary Objectives:

1. Estimate the proportion of transmission events that occur during acute and early HIV infection (AEHI).
2. Identify demographic, clinical, and epidemiologic factors that contribute to HIV transmission.
3. Estimate the proportion of transmission events that occurred within vs. outside of the HPTN 071 study communities.

Secondary Objectives:

1. Evaluate the prevalence and spread of antiretroviral drug resistance in the study communities.
2. Use coalescent modeling to estimate current and past trends in population-level HIV incidence.
3. Contribute samples to the Gates-funded Phylogenetic Consortium (known as PANGEA-HIV: Phylogenetics and Networks for Generalized HIV Epidemics in Africa).

Study Sites: Nine communities in Zambia will participate in the phylogenetics ancillary study (Chimwemwe, Ndeke, Ngungu, Maramba, Dambwa, Shampande, Chifubu, Chipulukusu and Makululu)

Funding: Funding has been provided by the National Institutes of Health for Stage 1 of the ancillary study. Core funding for Stage 2 and initial phases Stages 3 and 4 has been provided by the Bill and Melinda Gates Foundation as part of the PANGEA-HIV Phylogenetic Consortium. Funds for completion of Stages 3 and 4 have not yet been identified.

Oversight: The ancillary protocol team recognizes the primary importance of delivering the aims of the main HPTN 071 study. Data collection in this study will only initiate after the main study protocol team (and in particular the study leadership team, the HPTN LC and the HPTN SDMC) have reached a consensus to proceed. Approval to initiate may be provided on a community by community basis.

HPTN 071-2: Phylogenetic Ancillary Study to HPTN 071 (PopART)

1. SUMMARY

HPTN 071 (PopART) is a cluster-randomized HIV prevention trial with an HIV incidence endpoint. The trial includes 21 communities (12 in Zambia; 9 in South Africa), each of which is defined as the catchment population of a government Health Care Facility (HCF). These communities have been arranged in matched triplets. Within each triplet, communities have been randomly allocated to one of three study arms: Arm A: full PopART intervention, including antiretroviral therapy (ART) for all HIV-infected individuals regardless of CD4 cell count; Arm B: full intervention but with ART delivered according to national guidelines; Arm C: standard of care. The main trial will determine the impact of the study interventions on HIV incidence. Even if one or both of the PopART interventions is successful at significantly reducing HIV incidence, new cases of HIV infection will occur in all three study arms.

This ancillary study will use phylogenetic methods to examine patterns of HIV transmission on a population level in selected HPTN 071 communities representing all three arms of the trial. This study will provide key insights into the source of new infections in these communities and will identify factors associated with HIV transmission. The challenges, in terms of logistics, numbers of samples, and complexity of data analysis, are considerable. However, the potential insights we may gain through this analysis are substantial. This information will help guide future studies and public health prevention programs that build on the results of the HPTN 071 trial to optimize community-based approaches for HIV prevention.

As described below, this ancillary study will use novel approaches to understand the impact of the HPTN 071 interventions in Zambia on HIV transmission, to understand why these interventions do not prevent all new HIV infections, and to identify key factors that fuel on-going transmission. We note that similar approaches are also planned in other trials evaluating ART for HIV prevention, such as the Botswana BCPP and the Africa Centre TasP trial.

The phylogenetic study has been designed to address topics that are hypothesized to affect the outcome of the trial, and also long-term projections for the success of the PopART interventions. These include the role of acute and early HIV infection (a period of raised infectiousness after infection) in transmission, the role of transmission due to partnerships occurring outside of the trial, the role of people of specific sex and age groups, and the role of transmitted drug resistance. The phylogenetic study will be the main source of information on these topics. Output from the phylogenetic study will inform the mathematical and economic model projections regarding the long-term impact of the PopART intervention.

2. BACKGROUND

HIV infection is usually initiated by one or a few viral variants. Over time, HIV diversity usually increases, producing a swarm of genetically related viral variants. The viral evolutionary trajectory of HIV in a single host can be viewed as similar to the global molecular evolution of RNA viruses causing acute infections, such as influenza [1]. Factors that contribute to the HIV diversification include large viral population size, high mutation rate, lack of proof-reading of viral polymerases, frequent recombination, short replication time, rapid viral turnover, and selective pressures. The degree of viral

diversity and the rate of viral diversification vary from individual to individual and from region to region within the viral genome [2]. HIV diversity can drop to low levels in advanced HIV disease and certain settings (e.g., continued use of a non-suppressive antiretroviral [ARV] drug regimen). Globally, the extensive genetic diversity of HIV is reflected in the large number of co-existing subtypes and circulating recombinant forms, particularly in sub-Saharan Africa. In the HPTN 071 communities, the great majority of infections are expected to be HIV-1 subtype C.

The genetic features of HIV in an infected individual are unique and serve as a “viral fingerprint”. Importantly, because genetic features of HIV are likely to be similar among individuals who are linked through transmission (e.g., a transmitter and infected partner), HIV sequences can be used to infer putative transmission linkages and thus study the dynamics of HIV spread in networks, communities, and populations. This type of analysis is usually performed using phylogenetic methods that generate “trees” showing the genetic relatedness and ancestry of HIV from different individuals.

HIV phylogenies tend to reflect the demographic and spatial history of transmission in a specific setting [3-6]. In population-level analyses, HIV phylogenetic inferences have been remarkably successful in describing the origin and global spread of HIV, even working from very limited sequence data [3-8]. To date most large-scale phylogenetic studies (i.e., studies that analyze many viral sequences) have focused on concentrated epidemics in low prevalence settings due to the greater availability of viral sequence data (e.g., data generated during drug resistance testing performed for clinical care). Recent studies of concentrated epidemics have demonstrated that viral polymerase gene (*pol*) sequences generated during drug resistance testing, when associated with individual-level demographic and clinical patient data, can be used to characterize HIV transmission dynamics on the population level [9-13]. The most comprehensive European drug resistance databases reach 45% sampling coverage of at least one viral sequence for each HIV-infected individual. These studies in concentrated HIV epidemics and their associated methodological developments provide a scientific precedent for the population-level phylogenetic study proposed here.

Use of phylogenetic analysis in HIV prevention trials is relatively novel. Building on earlier studies that used phylogenetics to examine individual transmission events [14-17], investigators recently used phylogenetic methods to analyze transmission linkages in HIV prevention trials that enrolled serodiscordant couples. The utility of phylogenetics in HIV prevention research was demonstrated most clearly in the HPTN 052 trial. HPTN 052 used ART to prevent transmission in serodiscordant couples by treating the HIV-infected individual in the couple, either upon enrollment or according to in-country guidelines [18]. Prior to un-blinding in the trial, 4 individuals became infected during the trial in the intervention arm, versus 35 in the control arm (hazard ratio = 0.11, 95% CI 0.04-0.32) [18]. However, phylogenetic analysis demonstrated that 3/4 of the newly-infected individuals in the intervention arm and 4/32 of the newly-infected individuals in the control arm were not likely to have been infected by their HIV-infected partner [19]. The hazard ratio for linked HIV infection from the HIV-infected index participant was 0.04 (95% CI 0.01-0.27), which provided a much higher estimate of ART efficacy for HIV prevention than was obtained without considering the results from the viral linkage studies [18]. A similar phylogenetic analysis was performed in the Partners in Prevention study [20]. These examples provide precedent for the use of phylogenetics in HIV prevention trials.

Phylogenetic analysis can also be used to identify risk factors associated with HIV transmission. So far, most work on HIV phylogenetics has been done in the context of concentrated epidemics; a particular

novelty of this study is the application of phylogenetics to quantify routes of transmission in a generalized high-incidence epidemic. In concentrated epidemics, separation of viral phylogenies into reliable clusters has demonstrated that national HIV epidemics are further divided into distinct transmission networks within high-risk populations [9-12, 21]. Simple tests of association have demonstrated that concentrated transmission networks are not necessarily separated spatially [22], but often according to risk and demographic factors [23-27]. Viral phylogenetic inference has also proven useful to quantify epidemiological determinants of onward transmission such as viral load, disease stage, CD4 cell count or presence/absence of sexually transmitted infections (STIs). In a well-studied population in Brighton, UK, Fisher et al. [28] identified the most likely transmitters in a well-characterized longitudinal cohort of men who have sex with men (MSM), and identified risk factors for transmission by regression analysis. The study population included 700 men with established HIV infection and 159 newly infected individuals. Of the 159, the likely transmitter was found for 41 (26%) of new infections, roughly in line with expectations based on the coverage of the study. Factors associated with transmission in multivariable analysis were: younger age (rate ratio per 5 years older: 0.68 [95% confidence interval (CI) 0.54–0.86]), higher viral load (rate ratio per log higher: 1.61 [95% CI 1.15–2.25]), recent infection (rate ratio: 3.88 [95% CI 1.76–8.55]) and recent STI (rate ratio: 5.32 [95% CI 2.51 – 11.29]). ART was highly protective in a univariate analysis (rate ratio: 0.14 [95% CI 0.07–0.27]). Individuals with recent infection accounted for 11 (27%) of the transmissions; other individuals who were undiagnosed at the estimated time of transmission accounted for 19 (46%) of transmissions; and individuals who discontinued ART accounted for 9 (22%) of transmissions. This study serves as an important template for how a phylogenetic study may work in HPTN 071, albeit at a much more limited scale and in a very different context. This approach to risk ratio estimation can be embedded into sophisticated Bayesian estimation frameworks that use all of the information in a phylogeny, not just the most likely transmission events [29].

Phylogenetic clustering analyses have also successfully documented external sources of concentrated HIV transmission networks. For example, injecting drug user populations were shown to seed multiple transmission networks among heterosexuals in Eastern Europe [30], and migrating or traveling populations were identified as introducing new sub-epidemics in particular surveillance regions [11, 25]. It is more difficult to quantify the impact of migration, especially when the analysis includes an insufficient number of background sequences from epidemiologically-relevant populations outside the main study area. A recent study addressed this issue by applying phylogenetic methods and concepts to a well-studied generalized HIV epidemic in Africa, the Rakai Community Cohort in Uganda, cross-validating the findings with data from behavioral surveys. By comparing individuals whose sequences did not cluster with sequences from other individuals who reported having partners who lived outside of the Rakai area, and combining both sources of information in a probabilistic model, it was estimated that approximately 2/3 of transmission events occurred outside of households, and that among extra-household transmissions, the majority occurred outside of the community [31]. Of note, the clusters that constitute the Rakai cohort are much smaller than those in HPTN 071; the generalizability of these findings is unknown.

Finally, it is important to note that HIV super-infection (the re-infection of someone with a new viral strain in addition to the original virus) can complicate the analysis of generalized epidemics [6, 32]; full-genome phylogenetic analyses may help address this problem [33]. In addition, use of whole or nearly whole genome deep sequences may improve the accuracy of phylogenetic analyses in settings where viral super-infection and resulting inter-strain recombination are common. Deep sequencing refers to the application of next generation sequencing (NGS) methods to characterize the viral population in a single

sample. Using this approach, fragments of HIV genomes from many viruses in the ‘swarm’ infecting a patient are sequenced; the diversity of those viruses is assessed as part of the ‘viral fingerprint’.

3. STUDY OBJECTIVES AND METHODS

3.1 General approach

This study will generate HIV viral sequences from a large number of individuals in nine selected HPTN 071 communities in rural Zambia. The large size of the communities surrounding Lusaka and the clinic burden at the HCFs in those communities mean that it will not be feasible to include these communities in this study.

These viral sequences will be linked to epidemiologic, clinical and laboratory data. HIV phylogenetics will be used to examine individual-level putative transmission events and transmission clusters between anonymized viral sequences [34]. In some cases, linkages may be weak or uncertain; however, when analyzed in aggregate in a probabilistic framework that averages over many possible links, these data can provide strong population-level evidence on the drivers of the HIV epidemic in the study communities. The relationships between the sequences will also be analyzed in the context of associated data to identify factors associated with HIV transmission. This analysis will require obtaining HIV sequences from individuals with prevalent HIV infection (i.e., those who were already infected at the start of the trial), and incident infection (i.e., those who acquire HIV infection during the trial). For this approach to be successful, the coverage of sequenced viruses (i.e., the proportion of the population contributing samples for analysis) must be very high. This is a new and rapidly developing field of phylogenetic research. A large consortium of researchers in the field, as well as ethicists and community representatives, are closely involved in the design of the proposed studies.

This study will use blood samples collected from two sources in nine communities in Zambia: (1) participants enrolled in the PC of the HPTN 071 trial, and (2) patients attending HCFs in the communities (see Schema and Section 6). A key feature of this ancillary study is that it will not require contact tracing, invading the privacy of study participants, or otherwise compromising the anonymity of study participants.

Inclusion of samples from the PC will mean that at least some of the samples studied will come from an unbiased set of individuals, including many incident cases. To obtain sufficient coverage, individuals from HCF will also be included in the study, such that a higher fraction of the infected population will be sampled. As there will be two sources of data available, this will allow biases to be assessed.

This study includes the following specific objectives, listed in the Schema and described in detail below.

PRIMARY OBJECTIVE 1: Estimate the proportion of transmission events that occur during acute and early HIV infection (AEHI).

Rationale:

AEHI refers to the period lasting ~1-6 months following HIV acquisition. During the start of this period, individuals often have high viral loads, which are associated with very high infectiousness [35].

Individuals with AEHI are often unaware of their HIV status, and hence do not modify their sexual practices. The duration of elevated infectiousness is uncertain, as is its contribution to net transmission [36-38]. Available estimates on the population-level proportion of HIV transmission events driven by AEHI vary widely from 5-90% [35]. In HPTN 071, HIV testing will be offered annually using routine HIV screening tests by community HIV care providers (CHiPs) in Arms A and B as part of the PopART intervention package, whilst arm C communities will offer HIV testing according to standard care. Even with perfect coverage within the intervention arms of the trial (A and B), most HIV-infected individuals identified during the study will not start ART until after they have passed through the AEHI period, and so transmission during AEHI will not in general be prevented. This is especially the case if infectiousness associated with AEHI lasts a very short period of time. The degree to which transmissions occur during AEHI could reduce the impact of the PopART interventions in Arms A and B is unknown [39]. By adapting phylogenetic approaches previously used to study concentrated HIV epidemics [28, 29, 40], this ancillary study will attempt to quantify the contribution of AEHI to HIV incidence in the three arms of the HPTN 071 trial in Zambia.

Approach:

A variety of approaches will be used to identify newly infected individuals who were likely to have had AEHI at the time of sample collection or shortly before. These approaches are described below.

PC participants:

In the PC, participants who are HIV infected at enrollment in the HPTN 071 trial will not have AEHI at subsequent study visits when samples are collected for this ancillary study (i.e., at PC12, PC24, and PC36); those participants will have been infected for at least 1-3 years at the time of sample collection. Those cases will be classified as prevalent infections. Samples for this ancillary study will be obtained from PC participants who acquire HIV infection during the trial; those participants will mostly have been infected for less than one year (except for those who have missed one or more visits prior to their diagnosis). These individuals will be classified as incident infections. It will be assumed that individuals will have spent 50% of the previous year in AEHI. If samples from PC0 are available and the participant gives consent for their use for this ancillary study, additional incident infections in the PC may be identified.

HCF participants: For these participants, the duration of HIV infection will not always be known. In this group, participants who are likely to have had AEHI may be identified using a multi-assay algorithm (MAA), similar to the approach that was developed by the HPTN Laboratory Center (LC) that was developed for identifying recently-infected individuals in epidemics that are predominantly subtype C [20]. The MAA used in that study included two serologic assays (the BED capture immunoassay [BED-CEIA] and the BioRad-Avidity assay), and two non-serologic biomarkers (CD4 cell count and HIV viral load). That MAA has a positive predictive value of 93.9% for identifying recently-infected individuals (infected <180 days) [20]. HPTN LC investigators are actively working on development and validation of alternate MAAs for use in subtype C populations; the MAA used for analysis this ancillary study will be identified closer to the time of testing. Note that HCF participants in this study will have CD4 cell count data; furthermore, sufficient plasma should be available for a subset of participants for testing with additional assays (e.g., serologic assays, HIV viral load).

Next-generation sequencing (NGS): In this study, NGS will be used for all sequence analyses, which will provide detailed information about the viral population in each plasma sample. These data will also be used to estimate the duration of HIV infection in study participants. As noted above, HIV diversity usually increases during the course of infection. While there is some variability in the timing and patterns of viral diversification, mutations tend to accumulate and become fixed in the viral population. Molecular clock methods can be used to date the time of the most common recent ancestor for individuals infected with HIV [10], and are implemented in the popular software BEAST [41]. This analysis can provide an earliest possible date of infection of individual in a transmission cluster, and therefore can be used to estimate how far back in time HIV infection occurred (with statistical uncertainty) [10]. This approach was recently used to determine the timing of HIV transmission in two couples in HPTN 052 [42]. In this project, NGS data will be used to quantify the diversity of viruses within an individual (see for example [43]). Preliminary data using this new technology support a model of accumulating diversity in early infection [44]. The accuracy of this method will be further validated using samples with known duration of HIV infection from the HPTN 071 PC and other sources. Our current plan is to sequence the entire HIV genome in four parts (see [43, 44]). This would maximize the resolution of the phylogenetic inference, which would allow molecular clock dating to be performed with more confidence than previous analyses that have typically focused on partial *pol* sequences generated during drug resistance testing [10]. Sequencing technology is developing rapidly. In the likely event that new sequencing methods become available before Stage 3 commences, the study team will consider these in comparison to the methods proposed here, and adopt new methods and technologies as appropriate.

After identifying individuals who are likely to have had AEHI at the time of sample collection using the methods described above, we will analyze the patterns of clustering in phylogenies. Co-clustering of newly-infected individuals (e.g., seroconverters in the PC, MAA-positive individuals attending HCFs) will count as contributions to transmission from AEHI. In contrast, clustering of prevalent cases with newly-infected individuals will count as contributions to transmission from individuals with established infection. Inferences will be based in a statistical framework that adjusts for biased sampling. Using very conservative methods, the study is powered to distinguish whether the proportion of transmissions attributable to individuals in AEHI is high (>40%) or low (<10%). Power calculations (see Appendix) indicate that for this objective to be successful, we will need high coverage (e.g., sequences from >25% of all HIV-infected individuals in a community). We will adjust our analyses to account for the fact that recent infections in the PC will be defined with very high sensitivity and specificity but with a window of one year, whereas recent infections in the HCF will be ascertained with the MAA with lower specificity and shorter infection window (6 months).

PRIMARY OBJECTIVE 2: Identify demographic, clinical, and epidemiologic factors that are associated with HIV transmission.

Rationale:

In the HPTN 071 PC, we will have demographic, clinical, and laboratory data for participants who acquire HIV infection during the trial. These data can be used to identify factors associated with HIV acquisition. In contrast, the trial will not have any data on the individuals who were likely to have been the source of HIV infection (i.e., transmitters, those driving the epidemic). In this objective, we will use phylogenetic methods to identify factors associated with HIV transmission. Consideration of these factors and their contribution to HIV transmission will extend methods of molecular epidemiology previously

used in concentrated epidemics [27, 28, 45, 46]. Results from these analyses may also help target interventions to selected sub-populations in future HIV prevention programs, improving their impact.

Approach:

In this analysis, individuals will be classified as likely transmitters (or not) using a combination of phylogenetic clustering, molecular epidemiology modeling, and incidence data. Based on this classification, we will estimate the relative risk of being a likely transmitter among all HIV-infected individuals in the sample associated with a range of potential risk factors for transmission following the approach used in the study of MSM in Brighton, UK [28]. For both PC and HCF participants, data will include age, gender, ART use (current and/or previous, based on self-report), HIV infection status (seroconverter vs. prevalent infection and date of HIV diagnosis, per Primary Objective 1), circumcision status for men, pregnancy status for women, viral load (if available) and partnership status (currently in a regular sexual partnership, married, etc.). For HCF participants, this will also include CD4 cell count. When feasible, these data may be collected via existing clinic data collection (e.g. chart notes and/or SmartCare). Data source will be clearly identified in site SOPs.

We will analyze the broader properties of transmission networks identified by the phylogenetic analysis. This analysis will establish the extent to which transmission is concentrated (i.e., in a few large clusters or hubs of transmission, as might arise due to onwards transmission during AEHI, or with transmission from a small number of very high-risk individuals), or diffuse (with transmission fragmented into transmission pairs and triplets). This analysis will establish the extent to which transmission is assortative with respect to age or other factors. In the case of clusters involving multiple members of PC (expected *a priori* to be few in number), this analysis may reveal the extent to which transmission is assortative with respect to other data covariates. This approach has been previously applied to molecular genetic surveillance data [11, 12, 27, 40, 49], and has established that in these settings, transmission networks are highly concentrated for MSM, but much more diffuse for heterosexuals, with high degrees of assortativity by demographic factors. Analyses of this kind have not been reported for generalized epidemics in sub-Saharan Africa.

A more sophisticated analysis will estimate the relative transmission hazards between risk groups based on a comprehensive multi-type branching process or coalescent model that integrates over all possible transmission chains that are consistent with the viral phylogeny [45, 50, 51]. This second approach is likely to be more robust to biased sampling and will make use of as much greater proportion of the data, but will also be much more complex to implement and interpret. In comparison, the benefit of the clustering analysis of the previous paragraph will be simplicity, but the method will be less powered due to misclassification and lower sample size. Together, the methods will provide a more complete analysis.

The HPTN 052 trial demonstrated that ART dramatically reduces HIV transmission to an uninfected sexual partner. However, individuals who report that they are on ART may still transmit to others if they are not taking ART (since self-report of ART use is not reliable [47, 48]), or if they are on ART but are not virally suppressed (e.g., due to poor adherence or HIV drug resistance). The initial analysis will be based on HCF samples from participants who report they are not on ART at the time of sample collection (see Section 6). HCF participants will initially be recruited from patients who come to the HCF for their first visit during the data collection period of this ancillary study (including both new patients and patients already in care) and report that they are not currently on ART. Sampling may be expanded to include patients reporting that they are on ART subject to logistical feasibility to handle larger patient volumes.

PC samples will be collected from all PC members and will thus include from the outset some individuals who are on ART; viral sequences will be obtained for the subset of these participants who are on ART but are not virally suppressed (samples from persons who are virally suppressed will not have sufficient HIV RNA for amplification and sequencing based on current technologies). As technologies evolve, it may become possible to sequence very low concentrations of HIV RNA or integrated HIV proviral DNA from blood of patients on ART, however the power calculations assume that such samples will not be available. These data (samples from individuals on ART, current ART use, ART initiation date) would then be used to estimate, in addition to the other proportions discussed above, the proportion of new infections transmitted from individuals who report that they are on ART, but are not virally suppressed at the time of sample collection.

PRIMARY OBJECTIVE 3: Estimate the proportion of transmission events that occurred within vs. outside of the HPTN 071 study communities.

Rationale:

Individuals in the HPTN 071 communities may acquire HIV infection from an individual who resides in the same community (same study arm); in a different community within the same matched triplet (different study arm); in a different matched triplet (same or different study arm); or outside of the trial communities (no PopART intervention). Delivery of interventions, such as ART, would not be expected to prevent HIV acquisition if HIV-uninfected individuals have partnerships with HIV-infected individuals from outside the PopART communities. This may lead the HPTN 071 trial to under-estimate the potential impact of the PopART intervention were they to be applied at a national level, where such trial contamination effects may be absent. The proposed study will attempt to quantify the relative contribution of these various sources of new infections. The model predictions and power analyses for the primary endpoint in HPTN 071 are based on the assumption that 5% of sex acts will occur with outside partners. However, community-specific estimates of cross-community transmissions are not available for the HPTN 071 communities. Furthermore, it is not clear how transmissions from outside of the community will be impacted by community size and other geographical factors (e.g., rural vs. metropolitan area).

Approach:

This study will estimate the proportion of incident HIV infections in each study community that can be assumed to be from the various community types as described above. The patterns of phylogenetic clustering in different communities in HPTN 071 will be compared amongst each other and to communities in other studies (e.g., in the Gates consortium, see Secondary Objective 3). Simulation studies based on modeling the transmission and evolution of viruses in the HPTN 071 communities will also be performed. Fitting joint transmission-evolution models to genetic data is possible, but is technically challenging and is extremely computationally intensive [52]. In cases where communities are geographically close or adjacent (e.g. the Southern Province triplets, see Figure 1), we will attempt to establish the proportion of transmission clusters that cross community boundaries. Results from these analyses may help quantify the potential dilution of the effect of the PopART intervention on HIV incidence within that particular community. This will be a particularly challenging analysis, with no precedent to demonstrate feasibility. However, it aims to estimate quantities that are particularly desirable for interpretation of the trial – this is a ‘high-risk high-reward’ objective.

SECONDARY OBJECTIVE 1: Evaluate the prevalence and spread of antiretroviral drug resistance in the study communities.

Rationale:

In the main HPTN 071 trial, HIV drug resistance will be analyzed in a subset of PC participants using HIV genotyping assays that are validated for clinical use (pending funding). The HIV sequences generated for the proposed study (from both PC and HCF samples) will include the *pol* gene. Those sequences can be analyzed for the presence of ARV drug resistance mutations, providing additional data on the prevalence of drug resistance in a larger portion of individuals in the study communities. If drug resistance testing is performed for individuals in the main HPTN 071 trial and those individuals are also part of this ancillary study, the resistance test results obtained using the different testing approaches could be compared (e.g., clinically-validated drug resistance testing based on population sequencing [main trial] vs. research methods using NGS [ancillary study]). Sequencing will be completed on data that is anonymized, with samples labelled using a code generated by an encryption key kept in a secure location separate from all research activities. In this study, samples will be processed and analyzed in non-clinical research laboratories; analysis of sequence data will also be performed in non-clinical research settings. While the sequences generated in this study may provide some information about the antiretroviral drug resistance, the methods used to generate this information are not designed or validated for clinical use. For these reasons, results obtained from sequence analysis will not be returned to study sites, PC or HCF participants, or their health care providers where it might be used for clinical decision making (e.g., selection of antiretroviral drug regimens for HIV treatment).

Approach:

Drug resistance mutations will be identified based on the assembled consensus sequence for each sample, which will be compared to reference *pol* sequences (e.g., from the Los Alamos HIV database) and to well-documented drug resistance mutations (e.g., using the Stanford HIV drug resistance database). Minority variants will also be detected if they are present above a threshold of 1% of the viral quasi-species.

SECONDARY OBJECTIVE 2: Use coalescent modeling to estimate current and past trends in population-level HIV incidence.

Rationale:

Phylogenetic methods have proved popular for estimating changes in population-level incidence over time, an approach termed phylodynamics [4, 7, 21, 29, 53]. Here, demographic models are coupled to evolutionary models in coalescent or branching process models, and estimation is implemented in the popular package BEAST [54]. We will perform phylodynamic analyses of all data collected in this ancillary study, which will provide some independent estimates of population-level incidence dynamics in the HPTN 071 communities, both during the trial and in the preceding years leading up to the trial [4, 29, 53]. These analyses may provide insight into the dynamics of the HIV epidemic in the study communities prior to the trial, and the impact of the study interventions on the primary endpoint of HIV incidence. We note that such estimates are typically accompanied by very wide statistical uncertainty, and tend to be more accurate in the distant than recent past [55], though BEAST analyses have not typically been performed on data sets as large or comprehensive as we propose to collect here.

SECONDARY OBJECTIVE 3: Contribute samples to the Gates phylogenetic PANGEA-HIV consortium.

This ancillary study proposes to perform cutting-edge science at a scale that has not been previously attempted in a challenging, high HIV incidence setting. To support this activity, a consortium of scientists interested in large-scale viral phylogenetics in sub-Saharan Africa has been brought together and is funded by the Bill and Melinda Gates Foundation. The consortium brings together groups responsible for long-term cohorts (Africa Centre, KZN, South Africa; Rakai, Uganda; MRC Unit Uganda) and those engaged in HIV prevention trials (Botswana CPP; HPTN 071). HPTN 071 participation in this consortium will provide the following benefits:

- Resources for viral sequencing.
- Access to shared experience and expertise in all aspects of this project.
- Precedent for conduct of viral sequencing and phylogenetic analysis performed without compromising individual anonymity.
- Access to comparable viral sequences from outside of the trial area, which will aid the interpretation of phylogenies generated from HPTN 071 data.

Collaboration with the Gates-funded PANGEA-HIV Phylogenetics Consortium will enhance the chance of successful delivery of this ambitious project. Approximately 2,000 samples collected at HCF sites will be provided to the consortium for analysis. These will be shared under a legally binding Material Transfer Agreement that specifically precludes any sharing of identifying data, so that from the perspective of the PANGEA-HIV Phylogenetics Consortium this will be fully anonymized data.

These will be submitted from 2015 onwards during the course of the trial, with basic demographic data including age, treatment status, and CD4 count, but without information that would compromise the trial, or allow any linkage back to those individuals. Sufficient plasma will be available to support the work in this ancillary study as well as work performed by the PANGEA-HIV consortium (see Section 7). Sequence data will be generated using a laboratory and bioinformatics pipeline identical to the one proposed here. Sequence data will be returned to the PopART team and will contribute to the delivery of this ancillary study.

4. STAGES OF THE ANCILLARY STUDY

The project is divided into four main components (Stages 1-4):

- **Stage 1: Sample and data collection.** This stage of the study includes collection of blood samples from individuals enrolled in the PC and individuals attending HCFs. Individual consent will be obtained from PC participants as well as individuals attending HCFs to participate in the ancillary study. This stage of the study will also include processing and storage of plasma specimens from the HCFs; processing and storage of plasma specimens from the PC will be done as part of the main HPTN 071 trial. Epidemiological and clinical data from the PC will be obtained from the main HPTN 071 trial data set. For the HCF samples, these data will be collected from clinic records/forms when feasible as well as a short survey administered in the

clinic. There is no new blood sampling for the PC or HCF, and no change to the lab procedures for the PC.

- **Stage 2: Laboratory preparation and training.** This stage of the study involves identifying candidate laboratories to perform work needed for the ancillary study; identifying HCFs suitable for pilot work; preparing detailed Standard Operating Procedures (SOPs) for all stages of the laboratory work; obtaining the equipment and other resources needed by the participating laboratories; training laboratory personnel on the methods needed for sample preparation and analysis; and completing pilot studies using HCF samples to ensure that the laboratories can perform all required activities at a high level. This will include Quality Assurance/Quality Control (QA/QC) measures designed to reveal any testing problems, such as sample mix-ups, sample cross-contamination, RNA degradation, amplification or sequencing failure, data mix-ups, etc. This work will be performed in the context of the PANGEA-HIV Phylogenetics Consortium. However, additional QA/QC support may be required to ensure successful execution of the work in the ancillary study.
- **Stage 3: Analysis of samples to obtain HIV sequence data.** This stage of the study involves the analysis of the stored PC and HCF samples. This stage of the study will include the following activities: preparation of PC specimens; shipment of PC and HCF specimens to the appropriate laboratories (identified in Stage 2), sample processing (e.g., viral isolation, RNA extraction, reverse transcription, DNA amplification), and DNA sequencing.
- **Stage 4: Phylogenetic and statistical data analysis:** This stage of the study includes assembly of viral sequences, creation and curation of the final datasets, reconstruction of the viral phylogeny, and statistical analysis of the population-level determinants of onward transmission under appropriate mathematical models.

As noted in the Schema, Stage 1 is currently funded by the National Institutes of Health. Stage 2 and the initial phases of Stages 3 and 4 activities are funded through the Gates PANGEA-HIV Phylogenetics Consortium. Additional funding will be sought for completion of Stages 3 and 4.

5. STUDY POPULATION

As noted in Section 1, the HPTN 071 trial comprises 21 communities (12 spread across Zambia and 9 in Western Cape, South Africa), each of which is defined as the catchment population of a government HCF. The matching of the seven communities in triplets and arm randomization is described in the main HPTN 071 protocol. The impact of the intervention packages (delivered in Arms A and B) on HIV incidence will be assessed by enrolling and following a random sample of 2,500 adults in each community (52,500 participants total, the PC). Individuals enrolled in the PC will receive four visits: at enrollment (referred to as PC0), and 12, 24, and 36 months after enrollment (referred to as PC12, PC24, and PC36).

As noted above, the study team and potential sponsors have concluded that it is not logistically feasible or affordable to conduct this ancillary study in all 21 of the HPTN 071 communities. Therefore, this ancillary study will be carried out in nine PopART communities in Zambia (see Table 1 and Figure 1 below).

Table 1: Characteristics of HPTN 071 communities.

Zambian communities selected for this ancillary study are shaded.

Community*	Country	HIV prevalence (%)	ART coverage** (%)	Population	Proportion adult *** (%)	Proportion of adults who are PC participants(%)	Triplet	Selected for this ancillary study
Chimwemwe	Zambia	16	24	42898	53	11.7	1	Yes
Ndeke	Zambia	13	30	33297	53	15.0	1	Yes
Ngungu	Zambia	17	17	38081	53	13.1	1	Yes
Maramba	Zambia	17	29	55011	53	9.1	4	Yes
Dambwa	Zambia	24	24	31629	53	15.8	4	Yes
Shampande	Zambia	18	37	41615	53	12.0	4	Yes
Chifubu	Zambia	18	32	60222	53	8.3	2	Yes
Chipulukusu	Zambia	16	20	45234	53	11.1	2	Yes
Makululu	Zambia	20	30	34623	53	14.4	2	Yes
Chawama	Zambia	16	15	129221	53	3.9	3	No
Chipata	Zambia	15	23	166251	53	3.0	3	No
Kanyama	Zambia	17	27	124284	53	4.0	3	No

*Communities will not be named in any publications. Randomization arm is not shown here for reasons of confidentiality.

**Defined as the proportion of HIV-infected individuals on ART.

***Source <http://www.un.org/esa/population/>

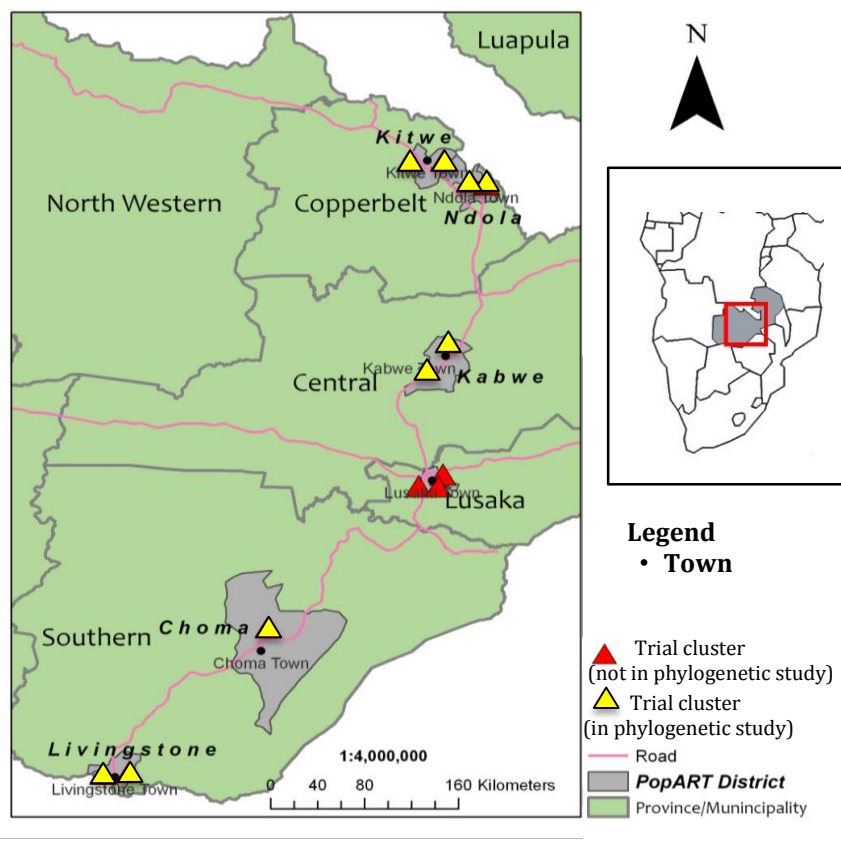


Figure 1: Map of communities (triplets) participating in HPTN 071 (all triangles), distinguishing between those communities in Zambia selected for participation in the phylogenetic study (yellow triangles) and those not selected for the phylogenetic study (red triangles).

6. SAMPLING AND DATA COLLECTION

6.1. Sample collection

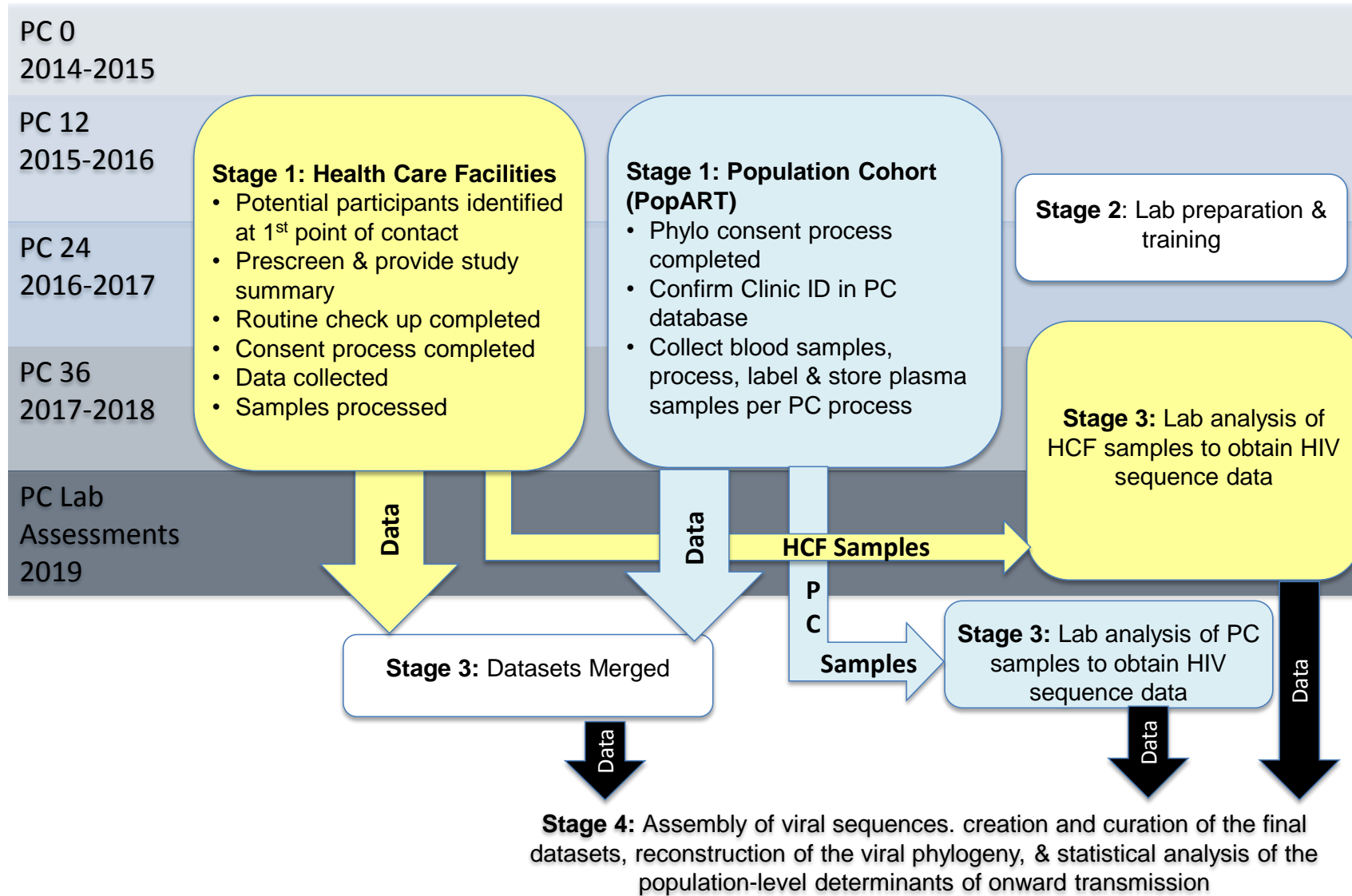
Stage 1 of the ancillary study will involve collection of samples and data from the PC and HCFs, as described above. The timeline and sampling strategy is shown in Figure 2.

Samples from PC participants for the phylogenetic ancillary study will be obtained from blood collected as part of HPTN 071 PC activities; no extra blood will be collected for the phylogenetic ancillary study. The ancillary study will be based primarily on analysis of samples collected at the PC12, PC24, and PC36 visits. Samples from the Enrollment visit (PC0, baseline) may be included, if enough plasma remains after laboratory assessments for the main HPTN 071 have been completed. Power calculations have been done under the assumption that such baseline samples will not be available; their inclusion would only increase the power of the study.

Sampling from the HCFs will begin at approximately the same time as sample collection begins from PC12 and will continue until the end of the PopART study for a duration of three years. Research and laboratory staff from ZAMBART will provide support to the HCFs for obtaining informed consent, data

collection and sample processing. One sample per individual will be included in the phylogenetic analysis. In some cases, individuals will be sampled in both the PC and HCF; procedures are in place to identify those individuals, primarily through the clinical patient identifier. Once the databases are merged, duplicate participants across the PC and the HCF will be identified by comparing Clinic ID numbers and the sample collected at the earliest date will be selected for sequencing. PC samples from all HIV-infected study participants who consent will be considered for inclusion in this ancillary study (see Section 6.3). The HPTN LC and HPTN SDMC will identify and release samples for this study based on pre-set criteria described below. Plasma samples from both the PC and HCFs will be labeled, tracked, and stored using the Laboratory Data Management System (LDMS).

Figure 2. Timeline, sample and data collection.



6.2. Target coverage and power calculations

There is no established methodology for performing power calculations for phylogenetic analysis. The study investigators have developed some methodology, based on the probability of inferring a transmission event involving a chronically-infected individual as an index case. If p_c denotes the probability that the chronic individual is included in the study, and p_i denotes the probability that the incident case is included in the study, then the proportion of transmission pairs included in the study is the product of the two, i.e. $p_c \times p_i$. (N.b. these numbers need not be equal or independent, but are a function of the overall incidence and take-up of phylogenetics in both PC and HCF.) If coverage of the phylogenetic study is 10 % ($=p_c=p_i$), then the proportion of pairs that is sampled is only 1% - if coverage is 30%, the coverage of pairs increases to 9% of all transmission pairs. Simple power calculations then indicate that for example, to achieve 90% power to differentiate between a scenario with very high levels of transmission attributable to index cases in acute infection (40% as in [56]) from a scenario with low levels (10% as in [37]) requires coverage greater than 25%. We thus propose a target of at least 30% of both chronically-infected and newly-infected individuals. Newly infected means infected during the trial, information which is known for members of the PC and inferred by MAA for samples obtained from the HCF.

6.3. Population cohort (PC) sampling

In the PC, participants in Zambia will be invited at the 12-month follow-up visit to participate in this ancillary study. If they agree to participate, a separate informed consent will be obtained (see Appendix 1A). If the participant consents to participate, they will also be consenting to use of samples at their prior (PC0) and latter visits. Blood samples will be collected from all PC participants as part of the main HPTN 071 trial; no additional samples will be collected for this study. All plasma samples from PC participants will be labeled with a participant id number to protect their identity. Data for many covariates will be obtained through the questionnaires administered at each PC visit. A detailed description of the questionnaire can be found in the main HPTN 071 protocol. Data that will be retained for PC participants for this ancillary study include:

- PC identifier (PTID)
- Clinical patient identifier (varies by laboratory and setting), for cross-referencing with HCF database. PC participants will be asked their unique patient identifier, but this is voluntary, so these data may be missing in some cases. This will be used to identify PC participants who also submit eligible samples to HCF (i.e. duplicate samples).
- Study community
- Year of birth
- Sex
- Current and previous use of ART (self-report)
- Date of sample collection
- Self-reported circumcision status (if male)
- Self-reported pregnancy status (if female)
- Partnership status (currently in a regular sexual partnership, married, etc.)
- HIV status (determined by the HPTN LC)

- Date of first positive HIV test (determined by the HPTN LC and HPTN SDMC)
- Viral load (performed retrospectively at the HPTN LC), if available
- Linked PC questionnaire data.

The HPTN LC and HPTN SDMC will identify and release PC samples for this study, using the following criteria:

- The PC participant is HIV infected at the relevant study visit.
- The PC participant provided informed consent for participation in the ancillary study.
- Sufficient plasma remains after laboratory testing for the main HPTN 071 trial has been completed (including quality assurance testing and testing for primary and secondary laboratory assessments).

6.4. Health care facility (HCF) sampling

The PC will not provide high enough coverage of communities to discover likely transmission links. Estimated coverage (proportion of adults in the community participating in the PC) ranges from 8% to 15.8% (see Table 1) in the communities selected for this study. In addition, the PC includes only one participant per household, which means that in-household transmission events cannot be inferred by analyzing PC samples. To meet the goal of achieving at least 30% coverage for phylogenetic analysis, additional samples will be obtained from individuals attending the HCFs in the relevant communities. These samples will be prepared using residual blood from samples sent from HCFs to laboratories for routine CD4 cell count testing.

Blood samples (4-5 mL) are routinely collected for CD4 cell count testing for all HIV-infected adults attending HCFs at biannual visits. Only 100 uL is required for CD4 quantification; the excess blood is then usually discarded. For patients attending the HCF, once CD4 cell count testing has been completed, the remaining whole blood sample for those who consent to participate will be processed to generate a plasma sample, which will be stored for the phylogenetic ancillary study. We will aim to process samples on the same day as they are collected.

Collection of samples will initially include only those patients who are not receiving ART at the time of the visit (based on clinic data), since samples from patients on ART using current sequencing technology usually have viral loads that are too low for amplification and sequencing. Participants will be recruited from HCF patients on their first visit to the facility during the period of this study, or for patients who discontinue ART, on their first visit to the HCF after discontinuation of ART. These will include newly diagnosed ART naïve individuals, patients registered in pre-ART care, as well as some individuals who have interrupted treatment. Collection of additional samples from patients who are on ART will be considered if the study team deem this feasible and that collection, processing and storage of the additional HCF samples will not compromise other activities in the main HPTN 071 study; this change will only be considered after interim review of key activities in the ancillary study and main HPTN 071 study by the whole study team. These samples are less likely to be from viremic patients, and so it will be harder to recover and sequence virus. Nonetheless they could add scientific value to the study, since some patients will be viremic (e.g., due to poor adherence, incomplete suppression, ART resistance).

Furthermore, advancing technology may allow the sequencing of viral RNA or proviral DNA from samples with lower viral loads in future; it will be some years before this study proceeds to Stage 3 (laboratory work). Power calculations below assume that only samples from HCF patients who are not on ART will be used.

In addition, HCF patients who agree to participate in this study will complete a short interview. The information collected from this interview and (when feasible) from existing clinic data will include:

- Phylo identifier (PTID)
- Clinical patient identifier (varies by laboratory and setting), for cross-referencing with PC database. This will be used to identify PC participants who also submit eligible samples to HCF (i.e. duplicate samples).
- Self-report participation in the PC (Yes/No)
- Lives in study community (Yes/No)
- Year of birth
- Sex
- Current and previous use of ART (self-report)
- Date of sample collection
- Self-reported Circumcision status (if male)
- Self-reported pregnancy status (if female)
- Partnership status (currently in a regular sexual partnership, married, etc.)
- CD4 cell count
- Date of first positive HIV test
- Viral load (if sufficient plasma and funds are available)

These data are needed for each visit (anticipated to be about every six months) of the participant to the HCF during the trial period, not just the visit where consent is obtained, to study correlations between the whole clinical history of the patient and the likelihood of being in a transmission cluster. For example, we would investigate whether discontinuing ART or failing to return for follow-up visits at the HCF is a risk factor for transmitting.

The sample will be labeled with a unique participant ID (PTID) and stored. When PC samples are available), the HCF and PC databases will be merged to identify any duplicate samples (i.e., samples that are likely to be from the same person). In some cases, several samples may be available for one person. These samples may be sequenced and used to verify the calibration of the within-host molecular clock used in analyses.

6.5. Estimated number of samples collected and identified for sequencing

The mathematical model described and used for the HPTN 071 protocol was further used here to predict the number of blood samples that would be stored at each HCF after CD4 cell count testing; these estimates assumed that individuals in pre-ART care would have a CD4 test once a year. These assumptions were used to estimate the number of samples from patients who consent that would be

collected at the HCFs for this study, and also to estimate the amount of data entry that will be needed when patients visit the HCFs during the study period. A summary of key assumptions made in the model and of consent rates in the HCF and PC is given in Table 2 below. We assumed high rates of consent for the PC and HCF as no additional bloods are being drawn for this study. Note that these estimates do not include PC0 samples or samples from HCF patients on ART.

The number of plasma samples to be stored was predicted based on the mathematical model (see Table 2 of main protocol, using central targets in Arms A and B and the best estimate model in Arm C).

Table 2: Assumptions about coverage in HCF and PC.

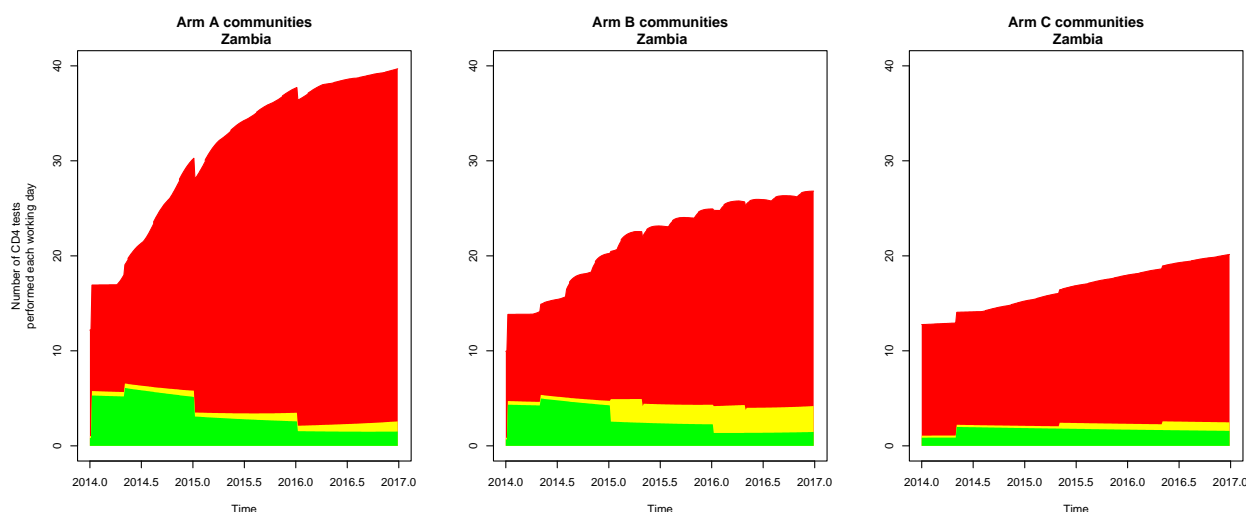
HCF: Proportion of untreated HIV+ individuals who link to care (Arms A & B) (%)*	HCF & PC: Proportion (%) of HIV+ individuals who are on ART at start of trial, by CD4 count category: CD4>350, 200<CD4<350, CD4<200)**	PC: Proportion of PC participants consenting to participate in the phylogenetics study (%)	HCF: Proportion of HCF participants consenting to participate in the phylogenetics study (%)	HCF & PC: Proportion of samples successfully stored for testing (Stage 3) (%)	HCF & PC: Proportion of samples successfully sequenced (Stage 3) (%)	HCF& PC: Contamination (Proportion of sex acts which are between one individual in and one individual out of the community) (%)
70	Zambia: 0, 25, 51	90	80	75	90	5

*Central target as in the main HPTN071 protocol.

** Percentages increase during the trial and were estimated as for the main HPTN 071 protocol.

The number of CD4 tests performed each working day at the HCFs for patients not on ART (broken down into first CD4 tests vs. repeated CD4 tests, pre-ART initiation), and CD4 tests amongst those already on ART, are shown in Figure 3. Volumes of CD4 tests are shown for information. The yellow and green areas indicate number of samples stored if only samples not on ART are stored, as initially planned. The red area shows the amount of data retrieved.

Figure 3: Average predicted number of CD4 tests performed each working day in each community HCF. The green zones show first-time CD4 tests; the yellow zones show repeated CD4 tests amongst individuals not yet on ART. The red zones show CD4 tests amongst individuals on ART. Abrupt changes in the predicted number of CD4 tests arise because model parameters are updated annually. The first noticeable increase in spring 2014 is a result of updated WHO guidelines.



The corresponding total numbers of plasma samples from the HCFs that must be stored over the trial period are shown in Table 3 (excluding any samples for those in the PC). The column ‘First CD4 tests’ corresponds to unique individuals at the HCFs providing their first sample for CD4 testing.

Table 3: Predicted total number of blood samples from unique individuals getting CD4 testing at HCFs in years 2, 3, and 4 of the trial, excluding individuals from the PC.

Country*	Arm*	CD4 tests performed in years 2, 3, 4 of the trial (total #)	First CD4 tests amongst untreated cases (in parentheses – amongst incident cases)	First CD4 tests amongst cases on ART only included in study if considered operationally feasible**
Zambia	A	87295	5581 (900)	6971
Zambia	B	59467	6436 (1016)	4551
Zambia	C	44504	4450 (205)	3926
Total	-	191266	16467 (2121)	15448

* To avoid revealing randomization, numbers are aggregated across the three communities in each arm). This applies to Tables 3, 4 and 5.

** These samples will only be included in the sample if considered operationally feasible . They are not considered in the power calculations, or in sample sizes listed in tables below.

In the same way, the mathematical model used to derive targets for the HPTN 071 trial was further used to predict the total number of HIV-infected individuals in the PC who will provide at least one blood sample (see Table 4). Coverage assumptions are detailed in Table 2.

Table 4. Predicted total number of HIV-infected individuals in the PC within each arm.

Country	Arm	Prevalent infections (persons infected at start of trial), #not on ART** (in parentheses: # on ART, # lost/not consenting)	Seroconverters total # over 3 years (in parentheses: # identified through PC12, PC24 & PC36)
Zambia	A	366 (507, 204)	141 (66, 42, 33)
Zambia	B	486 (393, 207)	177 (72, 54, 51)
Zambia	C	642 (258, 210)	246 (93, 81, 72)
Total	-	1494 (1158, 621)	564 (231, 177, 156)

**These estimates do not include samples collected at PC0.

Based on these estimates, we expect to collect approximately 16,450 first CD4 samples from HCFs and 2,050 samples from the selected PC communities included in this study. Of these 18,500 samples, we expect approximately 2,700 to be from incident cases (individuals infected during the trial).

The total number of plasma samples described in Tables 3 and 4 were then added to predict the number of sequences that would be obtained for prevalent and incident infections in each community. Table 2 details the assumptions used to estimate the reduction in sampling due to amplification and sequencing failure.

The numbers were further adjusted by three factors: (1) loss to follow up in the PC (as described in the main HPTN 071 protocol). We assume retention in the PC of 90% at 12 months, 80% at 24 months and 75% at 36 months, as in the main HPTN 071 trial. (2) laboratory ‘loss’ (e.g. due to labeling problems, insufficient plasma) assumed to affect 25% of samples, and (3) amplification or sequencing failure (e.g., due to low viral load or primer mismatch), assumed to affect an additional 10% of samples. Finally, it was (conservatively) assumed that all eligible members of the PC also provided samples to the HCF; the number of potentially available samples was reduced to account for duplicate samples. In this manner, we obtained predictions for the number of sequences that we would obtain from unique eligible individuals.

Table 5: Predicted total number of sequences from unique individuals from the PC (starting from PC12) and the HCFs.

Country	Arm	sequences of new infections identified in PC (total #)	sequences of new infections identified in HCF (total #)	sampling coverage (% of all new infections)	sequences of HIV+ individuals infected before trial start and identified in PC (total #)	sequences of HIV+ individuals with established infection identified in HCF (total #)	sampling coverage (% of all HIV+ infected before trial start)
Zambia	A	75	403	23	210	2281	27
Zambia	B	99	485	27	279	2642	37
Zambia	C	141	100	6	372	2069	23
Total	-	315	988	-	861	6992	-

These estimates of coverage are based on mathematical model projections of the number of individuals attending HCF at different stages of infection, and make conservative assumptions.

In summary, based on the estimations and calculations above, we expect that the study will yield the following sequence set:

Table 6: Estimated number of samples and sequences included in this ancillary study.

	Sequences from individuals with new infections	Sequences from individuals with established infections	Total sequences
PC participants	315	861	1176
HCF patients	988	6992	7980
TOTAL	1303	7853	9156

7. LABORATORY PREPARATION AND TRAINING

The laboratories that will perform testing for the ancillary study have not been identified. Stage 3 of this study (preparation of the study samples for sequencing) will not progress until sufficient funding and laboratory capacity is in place to ensure that such a large number of samples can be correctly processed. After site activation, initial activities will focus on sampling from HCF and submission through the existing sample pipeline of the PANGEA consortium (see below). For the first batches of samples, particular attention will be paid to developing processes that ensure that a large number of samples can be prepared and sequenced at each participating laboratory without sample mix-ups, cross-contamination, RNA degradation, etc.

A consortium has recently been funded by the Gates foundation for the study of phylogenetics in sub-Saharan Africa (The PANGEA-HIV consortium). PANGEA funding covers the period Nov 2013-Nov

2016 and represents a collaboration with the Wellcome Trust Africa Centre, Wellcome Trust Sanger Centre, Imperial College London and others. The consortium aims to generate approximately 20,000 viral sequences (whole viral genomes) including samples collected from HCFs in Zambia through this ancillary study. Laboratories will be identified, equipped with the equipment necessary for testing, and trained to conduct testing needed for the ancillary study. Prior to transferring samples, they will be labeled with an identification number which is not linked to their clinic ID.

The goals of the consortium, in addition to generating the viral sequence data from a number of well-studied populations, are to enhance collaboration between study teams. The PANGEA-HIV consortium will address a number of scientific questions, which include the following five broad questions:

1. What can be inferred about epidemic dynamics and sexual network characteristics from phylogenetic and self-reported epidemiologic data? What does that imply for control strategies?
2. How does NGS data (whole or near-whole genome sequences) improve the inference of transmission dynamics?
3. Should phylogenetic analysis form part of the standard tool kit of epidemic surveillance, and inform trial design?
4. What are the transmission dynamics of a generalized epidemic and how do they differ from those of a concentrated epidemic where data are already available?
5. What are, at the individual level, the characteristics of infectiousness? Can we identify individuals at greater risk of transmitting the virus, and should these be prioritized for frequent testing and immediate ART?

The following groups are expected to submit approximately equal numbers of samples to the consortium: Wellcome Trust Africa Centre (from DHS site and TasP trial), Botswana/Harvard Combination Prevention Programme (from Mochudi observational study and OGAC-funded treatment as prevention trial), MRC Uganda team (from multiple observational cohorts), the Rakai Health Services Program (from the main cohort study and a study of nearby fishing villages), the Malawi/UNC observational and intervention studies, and HPTN 071. The main aim of the consortium is to use phylogenetics to gain an improved understanding of the transmission of HIV in generalized epidemics in sub-Saharan Africa, and use this information to evaluate and improve methods for HIV prevention.

The executive committee of the PANGEA-HIV consortium includes investigators from the institutions listed above, the PI of the Zambian site and the PI and co-PI of HPTN 071. This committee will oversee all procedures and analyses from this ancillary study.

As several groups are already sending samples to PANGEA from a variety of sites, every effort will be made to learn lessons from early experiences of other groups. PopART members attend regular PANGEA teleconferences and meetings.

8. SAMPLE ANALYSIS

Funding has not yet been requested for analysis of samples to obtain HIV sequence data (stage 3), beyond the funding already available through the PANGEA-HIV consortium. Specific plans for sample analysis (e.g., RNA extraction, reverse transcription, amplification, sequencing) have not been finalized.

Based on estimates provided in Section 6, we estimate that this ancillary study will generate ~1200 PC sequences and ~8000 HCF sequences for phylogenetic analysis (~9200 total). For HCF samples where there is sufficient plasma available, a MAA may be used to identify samples from individuals who are likely to be recently infected (see Section 3). This part of the study will be subject to funding. As noted in Section 3, the HPTN LC has already validated a MAA for use in subtype C settings that has a 93.9% positive predictive value for identifying recent infections [20]. It is likely that an even more robust MAA will be identified by the time this work is performed. Therefore, the MAA used in this study (i.e., the specific assays and the assay cutoffs used to identify recent infections) will be selected closer to the time of analysis. The HPTN LC is now focusing its efforts on optimizing MAAs for identification of recent infections in subtype C settings. This work is also funded by an NIH-funded R01 study (Eshleman and Brookmeyer, co-PIs, R01-AI095068).

To manage a project this size, the HIV analysis (e.g., HIV RNA isolation, reverse transcription, amplification and DNA sequencing) will be performed in one or more laboratories which are likely to be outside of Zambia. The HPTN LC will provide guidance on Quality Assurance/Quality Control procedures for this work. The technology for NGS is rapidly evolving. The methods and sequencing platform to be used in this study will be determined before a pilot study is undertaken and may use a combination of population-based sequencing and NGS. NGS offers advantages in terms of phylogenetic resolution; this could involve amplification and sequencing of 1 to 4 amplicons, each covering approximately one quarter of the HIV genome. This protocol is described in [43], and has recently been adapted to the comparatively inexpensive and scalable Illumina MiSeq platform.

HPTN071 contribution to PANGEA-HIV consortium

We will contribute up to 2,000 samples to the PANGEA-HIV consortium, selected from samples collected from HCF. The samples collected for this consortium will include pre-ART samples and all individual identification will be removed and re-coded prior to transfer of samples to the sequencing laboratories. In the legally binding Material Transfer Agreement, it will be stated that the key for the codes cannot be shared with PANGEA, so that all PANGEA related investigations will be carried out on anonymized data. Sequence data will be returned to the PopART team, and incorporated into the main data for this ancillary study by using the key held by PopART investigators in Zambia.

9. ANALYSIS OF SEQUENCE DATA

Funding has not yet been requested for phylogenetic and statistic and data analysis (stage 4). The current plan is for the Wellcome Trust Sanger Institute team to take the lead in terms of developing robust and portable protocols for preparing RNA for sequencing (reverse transcription, amplification and library preparation), sequencing, and primary bioinformatics. The phylogenetic analysis will be led by the research team of the Principal Investigator (Professor Christophe Fraser) in close collaboration with all study investigators.

10. ETHICS CONSIDERATIONS

10.1 Summary of ethical considerations

A critical issue when designing any phylogenetics study is to protect individual participant identities and, insofar as possible consistent with expectations of the in-country public health authorities, community identities as well. Protecting participant identities is necessary in order to prevent stigmatisation of individuals and social harms that may result if potential infection linkages were revealed. The key goal of this protocol is to determine patterns of viral transmission at a population level, not to determine individual level viral transmissions. Similarly, shielding community identities also protects the communities from stigmatisation, but this goal must be weighed against the value of the data the study may provide for optimal planning and targeting of public health interventions based upon transmission patterns within that particular community. Therefore, although the study publications will not identify particular communities by name, the study team after close communication with the community advisory boards, may consult with provincial and/or national public health authorities about community-level results, in case those authorities can use community-specific results to plan or target specific community prevention and treatment interventions. Consultation with community representatives about these approaches to minimizing risk have begun and will continue throughout the trial period.

Strict measures will be in place to safeguard confidentiality of data. All laboratory specimens, reports, study data collection, process, and administrative forms will be identified by coded numbers only to maintain participant confidentiality. Personal identifiers (name, clinic identity numbers) will only be collected for (1) informed consent and (2) operational and logistical purposes. Personal identifiers will appear on paper or electronically on consent forms and other listings. These listings will NOT include any (sensitive) study information (including laboratory data). A unique study number will be used to link personal identifiers to study information. Personal identifiers and link logs which may be on paper will be stored in a locked cabinet. Electronically kept personal identifiers and link logs will be stored in separate datasets with password protection only accessible for designated staff (for computers and servers). Hand-held devices will also be password protected and personal identifiers will be stored in an encrypted format.

Participants' study information will not be released without the written permission of the participant, except as necessary for monitoring by the National Institute of Allergy and Infectious Diseases (NIAID) and/or its contractors, representatives of the HPTN LOC, SDMC, and/or LC, other government and regulatory authorities, and/or site IRBs/ECs. Datasets transferred to locations outside the study sites (e.g. for analyses, progress reports) will be stripped of any personal identifier and encrypted before transfer.

All electronic data will be stored in password protected database systems. Read and write authorization of data will depend on the designation of the staff member. A second layer of protection is hardware password protection on computers, servers and networks. Thirdly data transfer over wireless or mobile networks will use Virtual Private Networks or router protected dedicated internet protocol addresses.

10.2 Community Engagement

Community engagement in HIV prevention research requires strategic plans for establishing contact with trial communities, introducing research concepts, sensitizing communities and potential participants, etc. HIV prevention research involving phylogenetic analysis will require a more targeted approach to community engagement in order to educate and secure support from multiple levels of community key stakeholders [57].

Community engagement has been initiated during the development phase of this protocol and community insights regarding the study will continue throughout the process. In particular, Community Advisory Boards will be asked to provide input into consent documents and processes and also to provide their views regarding the steps that will be taken to protect both individual and community identity and to reduce the risk of stigmatization. Individual written consent will be obtained from PC participants (see Section 10.3) and HCF attendees (see Section 10.4). An extensive process of community consultation will be undertaken in the participating communities for the implementation of the phylogenetics study. This community consultation process will consist of two broad phases:

- Phase 1 entails informing a broad array of community health stakeholders about the Phylogenetics study, responding to their concerns and advice, and ensuring that they are aware of, understand, and support the study and documenting this process by writing up minutes of all stakeholder meetings that include names of people who attended the meeting where possible. The list of stakeholders representing each community may differ, but will typically include representatives from traditional, faith, cultural, and civic leadership, civil society, the Ministry/Department of Health at multiple levels (facility, sub-district, district, province), and community based entities. The identification of stakeholders will be done in collaboration with the HPTN 071 Community Working Group and draw on the already completed BBS Formative Research (Social Science) conducted prior to HPTN 071 implementation. Following this stakeholder consultation, representatives from each of the participating communities will be asked to hold open community meetings where announcements about the impending commencement of the study are made, community members will be given the opportunity to ask questions to the research staff. These meetings will be fully documented by the study team (including attendance, content of discussions, and resolution of community members' queries). Channels of communication will be opened (prior to any research activities taking place within that community).
- Phase 2 will include an ongoing process of informing community members of the existence of the ancillary study in their community. The three main mechanisms for informing community members are: (a) the distribution of informational materials to households of PC participants participating in the phylogenetics study and to the ART sections of participating HCFs, (b) the informed consent document that is presented to PC participants and HCF patients, and (c) the use of existing community consultation platforms, including those created by HPTN 071 (e.g., the Community Advisory Boards), and those that are a regular feature of community life (e.g., Community Based Organizations) and support group meetings in Zambia.

Community engagement will continue following the initiation of the phylogenetics study procedures and is not limited to 'informing community members of the study's existence' (per phase 2 above). Scientific

leadership will provide training to the community engagement team to facilitate this on-going process. Updates on the progress of the phylogenetics component of the study will be included in regular engagement with community stakeholders about progress of the study. Community input will be sought on relevant issues that arise during conduct of the study, and, as described below, if requests are made by health officials for community-identified data from this study. The protocol team will consult and seek support from relevant advisory groups, officials and others to use blood samples from HCFs in the phylogenetics study. The protocol team is committed to providing feedback from findings of the analyses in stages 3 and 4 to participating study communities. Further development of such dissemination plans will be compiled as part of further funding arrangements for these stages.

10.3 Samples from PC participants

All PC participants in the nine selected communities in Zambia, irrespective of HIV status, will be asked to provide a separate written informed consent for participation in the phylogenetic study and details about the study will be provided by the study team. In the case of HIV uninfected PC participants no further use of their stored sample from that visit will be undertaken. For those PC participants not available to be offered participation at the 12 month follow-up visit, consent to participate in the phylogenetics study may be solicited at the 24 and 36 month follow-up visits, if the study team determines this to be operationally feasible. Consent given at these visits will allow use of samples collected at earlier and later visits.

10.4 Samples from HCF patients

All HIV patients not on ART who are attending a participating HCF during the sample collection period will be asked to provide written informed consent for participation in the phylogenetic study. For those who consent, samples of blood routinely collected for CD4 cell count testing as part of standard clinical care will be stored for use in this ancillary study.

10.5 Community anonymization

Throughout this study, the names of the individual communities will not be used in public presentations or publications of the study results. In the study database, however, community-level data will be discernable. These community links/identifiers will be maintained to preserve the possibility that community-specific results might be used by public health authorities to plan and target valuable prevention and treatment efforts. It is possible, for example, that aggregate community-specific data on levels of ART drug resistance could be a valuable public health tool for making local health policy on preferred future planning for ART regimens. Similarly, community-specific findings about the role of newly-infected persons in transmitting HIV and the role of sexual mixing with partners outside of the study communities on community-specific transmissions may be important for public health planning and resource allocation. There is risk that if community-identified data are shared with anyone, including public health authorities, those data might inadvertently or inappropriately be re-shared, with a result of stigmatizing specific communities. To avoid this, whenever possible, data unidentified as to community source will be made available for use by health officials for public health work or research; data identified

as to specific community or communities will only be shared when necessary and when public health benefit for the community is expected to outweigh the risk of community stigma. When, in the course of the study, the researchers believe that they may have identified results that could have public health value specific to one community, the research team would disclose, at a high level, those results to public health officials who have jurisdiction over that community. If those public health officials request more specific data about their community, then the researchers would consult with community representatives before deciding to disclose those data to the requesting public health authorities. In making such a disclosure, the researchers would advise the public health officials receiving the information that there is risk of community stigmatization if the identified data are shared irresponsibly, and that the data should be kept as confidential as possible, even while using those data to fulfill meaningful public health aims.

11.0 SAFETY MONITORING AND SOCIAL HARM/ADVERSE EVENT REPORTING

11.1 Safety Reporting

Because this study includes no biomedical intervention or study product, standard adverse event reporting will not be undertaken and no adverse event data will be collected on case report forms for entry into the study database. However, in accordance with 45 CFR 46, unanticipated problems or serious adverse events that are judged to be related or possibly related to study participation will be documented and reported to the IRB/ECs according to their individual requirements and to the DAIDS Medical Officer and the Office of Human Research Protections (OHRP). This reporting will be performed according to the timelines and definitions included in pre-established written procedures, such as the study procedures manual and SOPs, and the guidelines provided at <http://www.hhs.gov/ohrp/policy/advevntguid.html>. Serious adverse events will not be reported to the DAIDS Regulatory Compliance Center (RSC).

In order to prevent adverse social events related to study participation, social harms will be monitored throughout the study. Social harms are any untoward social occurrences that happen to a participant as a result of their participation in the study. Examples include loss of employment, harassment by neighbors, shunned by family, rejection by partner, etc. Although social harms due to this study are expected to be negligible, they will be monitored closely throughout. Information on social harms will be actively solicited from participants at follow-up visits and recorded on case report forms and captured in the study database. Participants will also be encouraged to report any social harm on an ad hoc basis when it occurs. In the event that a participant reports social harm, every effort will be made by study staff to provide appropriate care and counseling to the participant, and/or referral to appropriate resources for the safety of the participant as needed. Social harms that are judged by the Investigator of Record to be serious or unexpected will be reported to the responsible site's IRB/EC at least annually, or according to their individual requirements. The nature and frequency of these social impact reports will be monitored by the protocol team on a regular basis.

12. ADMINISTRATIVE PROCEDURES

12.1 Protocol Registration

Prior to implementation of this protocol, and any subsequent full version amendments, each site must have the protocol and the protocol consent form(s) approved, as appropriate, by their local IRB/EC and any other applicable regulatory entity (RE). Upon receiving final approval, sites will submit all required protocol registration documents to the DAIDS Protocol Registration Office (DAIDS PRO) at the Regulatory Services Center (RSC). The DAIDS PRO will review the submitted protocol registration packet to ensure that all of the required documents have been received.

Site-specific informed consent forms (ICFs) WILL NOT be reviewed or approved by the DAIDS PRO, and sites will receive an Initial Registration Notification when the DAIDS PRO receives a complete registration packet. Receipt of an Initial Registration Notification indicates successful completion of the protocol registration process. Sites will not receive any additional notifications from the DAIDS PRO for the initial protocol registration. A copy of the Initial Registration Notification should be retained in the site's regulatory files.

Upon receiving final IRB/EC and any other applicable RE approval(s) for an amendment, sites should implement the amendment immediately. Sites are required to submit an amendment registration packet to the DAIDS PRO at the RSC. The DAIDS PRO will review the submitted protocol registration packet to ensure that all the required documents have been received. Site-specific ICF(s) WILL NOT be reviewed and approved by the DAIDS PRO and sites will receive an Amendment Registration Notification when the DAIDS PRO receives a complete registration packet. A copy of the Amendment Registration Notification should be retained in the site's regulatory files.

For additional information on the protocol registration process and specific documents required for initial and amendment registrations, refer to the current version of the DAIDS Protocol Registration Manual.

12.2 Study Activation

Pending successful protocol registration and submission of all required documents LOC (HPTN Leadership and Operations Center) staff will “activate” the site to begin study operations. Study implementation may not be initiated until a study activation notice is provided to the site.

12.3 Study Coordination

Study implementation will be directed by this protocol as well as the study procedures manual. The study procedures manual — which will contain reference copies of the *Requirements for Source Documentation in DAIDS Funded and/or Sponsored Clinical Trials*, as well as the DAIDS Manual for Expedited Reporting of Adverse Events to DAIDS, Version 2.0, dated January 2010 and the DAIDS Toxicity Tables — will outline procedures for conducting study visits; data and forms processing; AE assessment, management and reporting; and other study operations.

Study case report forms, electronic data capture tools, and other study instruments will be developed by the protocol team. The study data from all sources ultimately will be transferred to Imperial College

London for storage and analysis. Quality control reports and queries will be generated and distributed to the study sites on a routine schedule for verification and resolution.

Close coordination between protocol team members will be necessary to track study progress, respond to queries about proper study implementation, and address other issues in a timely manner. Rates of accrual, adherence, follow-up, and AE incidence will be monitored closely by the team as well as the HPTN Study Monitoring Committee. The Protocol Chair, DAIDS Medical Officer, Protocol Biostatistician, SDMC Project Manager, and CORE Protocol Specialist will address issues related to study eligibility and AE management and reporting as needed to assure consistent case management, documentation, and information-sharing across sites.

12.4 Study Monitoring

Study monitoring for this ancillary study will be performed in accordance with the agreed upon monitoring plan for the main study.

- Site investigators will self-monitor according to their Clinical Quality Management Plan.
- DAIDS reserves the right to audit the self-monitoring on an approximately annual basis as may be appropriate for this ancillary study.

12.5 Protocol Compliance

The study will be conducted in full compliance with the protocol. The protocol will not be amended without prior written approval by the Protocol Chair and NIAID Medical Officer. All protocol amendments must be submitted to and approved by the relevant IRB(s)/EC(s) and the DAIDS RSC prior to implementing the amendment.

12.6 Study Discontinuation and Participant Termination

The study may be discontinued at any time by NIAID, the HPTN, other government or regulatory authorities, and/or site IRBs/ECs. Participants may be terminated from the study if they are unwilling or unable to comply with required study procedures or if the study is stopped or cancelled. Participants may choose to discontinue study participation at any time and for any reason without penalty or loss of standard, non-study related services.

12.7 Investigator's Records

The study site investigator will maintain, and store in a secure manner, complete, accurate and current study records throughout the study. The investigator will retain all study records for at least three years after submission of the Clinical Trial Unit's (CTU) final Financial Status Report to DAIDS, which is due within 90 days after the end of the CTU's cooperative agreement with DAIDS, unless otherwise specified by DAIDS or the HPTN LOC. Study records include administrative documentation — including protocol registration documents and all reports and correspondence relating to the study — as well as documentation related to each participant screened for and/or enrolled in the study — including informed consent forms, locator forms, case report forms, notations of all contacts with the participant, and all other source documents.

12.8 Use of Information and Publications

Publication of the results of this study will be governed by the HPTN Manual of Operations and policies. Any presentation, abstract, or manuscript will be submitted by the Investigator to the HPTN Manuscript Review Committee for review prior to submission.

13. CONCLUSIONS

The HPTN 071 (PopART) trial is a highly innovative trial that will evaluate a combination prevention package for HIV prevention. The trial will determine the impact of the study interventions on population-level incidence and will help identify factors associated with HIV acquisition – both in standard of care and intervention communities.

This ancillary study will extend those findings by evaluating how the study interventions impact the spread of HIV in the study communities, and identifying factors associated with on-going HIV transmission when these interventions are implemented (i.e., factors associated with those who transmit HIV infection, as well as those who acquire HIV infection).

The main study will clearly identify the determinants of uptake of the interventions and quantify how these translate into population level incidence. The study also has many secondary outcomes that will determine individual responses to the intervention, and benefit accrued, both in terms of risk of infection and other health outcomes. Only this phylogenetic study, however, will address the question of the source of residual infections. This is critical for interpretation, extrapolation and generalization of the findings of the trial.

This phylogenetics ancillary study will be one of the first very large scale studies using state-of-the-art molecular technologies to analyze the impact of interventions for HIV prevention on a population level. The work proposed here will be performed on an ambitious scale. This is necessary to provide the intensive sampling coverage needed for robust phylogenetic analysis, and also to evaluate the generalizability of the findings across clusters that vary in location, country, proximity of communities, and other factors.

APPENDIX 1: SAMPLE INFORMED CONSENT FORMS

HPTN 071-2: PHYLOGENETICS ANCILLARY STUDY

APPENDIX 1A: PC PARTICIPANT INFORMATION AND CONSENT FORM

Title of Research Study: HPTN 071-2 Phylogenetics in HPTN 071: An ancillary study to “Population Effects of Antiretroviral Therapy to Reduce HIV Transmission (PopART): A cluster-randomized trial of the impact of a combination prevention package on population-level HIV incidence in Zambia and South Africa”

Protocol #: HPTN 071-2, Version 1.0, 15 January 2015
DAIDS ID: 11865

Sponsor: Division of AIDS, National Institute of Allergy and Infectious Diseases
U.S. National Institutes of Health

Investigator of Record: (insert name)

Research Site Address(es): (insert address)

Daytime telephone number(s): (insert number)

24-hour contact number(s): (insert number)

Participant Information and Consent Form

Please ask the study staff to explain any words or procedures that you do not clearly understand.

Please read this Participant Information and Consent Form and ask as many questions as needed. You should not sign this form if you have any questions that have not been answered to your satisfaction.

The purpose of this form is to give you information about an additional study, which is part of the main HPTN 071/PopART trial that you are already participating in. If you sign this form, you will be giving your permission to join this additional study. This form describes the purpose, procedures, benefits, and risks of this additional research study.

Your participation is voluntary

You should only join the additional study if you want to do so. You can stay in the main part of the PopART trial even if you choose not to join the new study. You are free to withdraw from either the additional study or the main PopART trial at any time in the future if you change your mind. Choosing not to take part in the additional study will not affect your participation in the main PopART trial or the health care services that you or your family usually receives.

Who is doing the study?

Researchers from Zambia AIDS Related Tuberculosis (ZAMBART) Project and the London School of Hygiene and Tropical Medicine will be working closely together on this trial. This study is being funded by the U.S. National Institutes of Health and the Bill and Melinda Gates Foundation.

The study has been approved by the Medical Ethics Committees of the University of Zambia, Stellenbosch University, the London School of Hygiene and Tropical Medicine and Imperial College London.

What is the purpose of the HIV Phylogenetic study?

The purpose of this study is to help researchers understand how HIV spreads in the community. We will look at the different types of virus found in blood samples of different people in the community who are living with HIV. In science we call this Phylogenetics.

By itself, the PopART study won't tell us *how* the study activities slow the spread of HIV (or if these activities don't work, why they didn't work). That is the purpose of the study you are now being asked to join, called the phylogenetics study. I will now explain more about the phylogenetics study.

The virus in one person with HIV is slightly different from the virus in another person. This study will use special laboratory tests to compare the different types of HIV in people in communities where PopART is being done. The purposes of this testing are to:

- Learn if the study has made a difference in the way HIV spreads in the community.
- Give us a good idea of when people first became infected with HIV.
- Help us understand whether people with HIV got their infections from people inside the community or outside the community
- Help us understand whether people with HIV in the community got their infections from people who have had HIV for a long time, or from people who have only been infected recently.
- Provide information about whether some HIV in the community is resistant to the drugs used to treat HIV infection.

What will happen during this study?

If you agree to the additional study, we would like to do some additional tests on the blood samples that the researchers are already collecting from you for the main PopART trial. You will not need to provide any additional blood samples or attend any additional visits. The researchers will perform extra laboratory tests for this study on the blood you are already providing as part of the main PopART trial (from previous visits, today and at remaining visits). From the sample of blood we take we will test to see if you are infected with HIV. The researchers will also access some information about your health and answers to the questions you gave in your participation in the main PopART trial.

The tests that we are doing on your blood will be not be done until after the POPArt study is finished which may not be for several years. These results will be research tests only and therefore will not be helpful to your health care providers so no results will be provided back to you.

There is no cost to you or payment provided for participating in this study.

Who else will be asked to participate in this study?

We will do this additional study in nine of the Zambian communities that are part of the main PopART trial. Everyone in these communities who is already a part of the Population Cohort will be asked to participate in this additional study.

In addition, if they give permission, we will also collect blood from patients who receive HIV care at the participating health care facilities (clinics) in each of the communities. We will not collect extra blood from these patients either. We will only use what is left over after routine health care tests are done. If you go to one of those participating clinics, you may be asked to participate again since staff there will not know that

you are part of the PopART trial. If this happens, you may choose to participate or not at the clinic, as you wish.

We expect approximately 9,200 people to participate in this study.

All blood samples collected as part of this trial will be shipped and securely stored outside of the country for study-related testing. Samples will be stored indefinitely and used only for studies related to HIV as we discussed above.

What are the potential benefits?

There are no direct benefits to you, other than the satisfaction you may get from adding to the knowledge gained from this important study. This knowledge may help the research teams and health authorities worldwide understand how the PopART trial affected the spread of HIV in the study communities. It will help us to understand patterns of new HIV infections and how they spread through communities. It will also tell us how certain types of HIV medications help control the different types of HIV which people can have which can also help medical professionals make decisions about what kinds of HIV medicines to provide in your community.

What are the possible risks or discomforts?

There are no foreseen increased risks to being part of this additional study. However, there is an unlikely possibility that others may find out that you are participating in this study. In the unlikely event that this happens, you may experience stigma or other personal problems with family, friends, and people in your community. We will do everything we can to protect your privacy.

How will my confidentiality and privacy be protected?

The main concern that people may have is privacy. People may worry that a researcher would be able to know that one specific person infected another specific person with HIV. The study team will make sure that this is very unlikely to happen. If you are infected, we will make sure that no one is able to connect any information we learn about your HIV to you. Any information we have that would identify the blood samples as yours, like your name, will be permanently removed from the samples before tests are done on them for the additional study. None of the researchers or anyone in your community will ever be able to know that the samples of blood are linked to you. The only information that will link to your blood sample will be a study identification number which will be put on the tubes containing the blood, but it will not be possible to work out your identity from this information. In addition, we will never publish the name of your community in any results.

People who may review the study records include: [insert name of site IRB/EC], local regulatory agencies, US National Institute of Allergy and Infectious (NIAID) contractors, US National Institutes of Health (NIH), study staff, and study monitors. Institutional Review Boards (IRBs) or Ethics Committees (ECs) are committees that watch over the safety and rights of research participants. Provincial and/or national public health officials may be given community-wide results but not individual results.

What happens if participating in this study injures me?

It is very unlikely that you could be injured as a result of participating in this study. Nothing that we will be asking of you should place you at risk for injury. However, if you are injured while participating in this study, you will be given immediate treatment for your injuries. You will not have to pay for this treatment. There is no program for compensation through the United States National Institutes of Health. However, in case of any research related injury, the London School of Hygiene and Tropical Medicine would provide compensation.

Are there any alternatives to participation?

If we become aware of other research studies happening in this community that may benefit you, we will tell you about them.

What are some reasons why I may be withdrawn from this activity without my consent?

You may be withdrawn from the study without your consent if this research study, or this part of the study, is stopped or canceled.

Persons to Contact for Problems or Questions

If you have any questions about your participation in this research study, your rights as a research participant, or if you feel that you have experienced a research-related injury, contact:

Site research staff: (site insert name of the investigator or other study staff)

Research site address (es): (site insert physical address of above)

Daytime telephone number(s): (site insert telephone number)

24-hour contact number(s): (site insert telephone number)

If you have any questions or concerns about your rights as a research participant or want to discuss a problem, get information or offer input, you may contact:

Independent Review Board/Ethics Committee:

(Site insert name or title of person on the IRB, EC or other organization appropriate for the site)

Address of Independent Review Board: (site insert physical address of above)

Daytime Telephone Number: (site insert telephone number of above)

Thank you for reading this information sheet. If you have any questions, please ask them now. The interviewer will be pleased to answer them. If you wish to take part, please read and sign the consent form. We will offer you a copy of this this information sheet and encourage you to please keep it in a safe place.

**PARTICIPANT’S STATEMENT OF CONSENT
PHYLOGENETIC ANCILLARY STUDY**

HPTN 071-2 Phylogenetic in HPTN 071: An ancillary study to “Population Effects of Antiretroviral Therapy to Reduce HIV Transmission (PopART): A cluster-randomized trial of the impact of a combination prevention package on population-level HIV incidence in Zambia and South Africa”

- I have been given sufficient time to consider whether to take part in this study.
- My taking part in this research study is voluntary. I may decide not to take part or to withdraw from the research study at any time without penalty or loss of benefits or treatment to which I am entitled.
- I understand the difference between the main PopART trial and the Phylogenetics study.
- The research study may be stopped at any time without my consent.
- I have had an opportunity to ask the study staff questions about this research study. My questions so far have been answered to my satisfaction.
- I have been told that my participation in the additional study involves additional testing of my blood samples that will already be collected for the main PopART trial.
- I have been told who else will be invited to participate in this study, and that some blood samples will come from people attending HIV care at participating health facilities (clinics).
- I have been told what the possible risks and benefits are from taking part in this research study. I will not benefit individually if I take part in this research study.
- I do not give up my legal rights by signing this form.
- I have been told that before any study related procedures begin, I will be asked to voluntarily sign this Participant Information and Consent Form.
- I will receive a signed and dated copy of this Participant Information and Consent Form.

If you have either read or have heard the information in this Participant Information and Consent Form, if all of your questions have been answered, and if you agree to take part in the study, please print and sign your name and write the date on the line below.

I voluntarily agree to take part in this research study.

Participant’s Name (print) Participant’s Signature or Thumbprint Date

I certify that the information provided was given in a language that was understandable to the participant.

Name of Study Staff Study Staff Signature Date
Conducting Consent (print)

Witness’ Name (print) Witness’ Signature and Date Date
(If necessary)

APENDIX 1B: HCF PARTICIPANT INFORMATION AND CONSENT FORM

PHYLOGENETIC ANCILLARY STUDY

Title of Research Study: HPTN 071-2 Phylogenetics in HPTN 071: An ancillary study to “Population Effects of Antiretroviral Therapy to Reduce HIV Transmission (PopART): A cluster-randomized trial of the impact of a combination prevention package on population-level HIV incidence in Zambia and South Africa”

Protocol #: HPTN 071-2, Version 1.0, 15 January 2015
DAIDS ID: 11865

Sponsor: Division of AIDS, National Institute of Allergy and Infectious Diseases
U.S. National Institutes of Health

Investigator of Record: (insert name)

Research Site Address(es): (insert address)

Daytime telephone number(s): (insert number)

24-hour contact number(s): (insert number)

Participant Information and Consent Form

You are invited to take part in a study looking at ways to reduce the number of new HIV infections in your community. Information about the study is supplied in this document. Please make sure that you understand everything described in this document. If you decide to participate, you will be asked to give written consent before you take part. If you sign this form, you will be giving your permission to take part in the study.

This form describes reasons for doing the study, how we will conduct the study and the benefits and risks of the study.

Your participation is voluntary

You do not have to take part in this study. If you decide today to take part in this research project, you may stop at any time without reducing or affecting any care that you receive at this clinic or benefits that you or your family will receive.

Who is doing the study?

Researchers from the Zambia AIDS Related Tuberculosis (ZAMBART) Project and the London School of Hygiene and Tropical Medicine and partners including Imperial College London, will work closely on this trial with colleagues from the HIV Prevention Trials Network (HPTN). This study is being funded by the U.S. National Institutes of Health and the Bill and Melinda Gates Foundation.

The study has been approved by the Medical Ethics Committees of the University of Zambia, Stellenbosch University, the London School of Hygiene and Tropical Medicine and the Imperial College London.

What is the purpose of the study?

The study you are being asked to join is a part of a larger study taking place in Zambia and South Africa called the “PopART” study. In the PopART study, we are testing whether we can reduce the number of new HIV infections in a community by providing a number of HIV-prevention activities all at once in that community. These activities include providing HIV testing in people’s homes, linking people who do have HIV to care at clinics, promoting circumcision for men who do not have HIV, providing condoms in the community, and other activities.

By itself, the PopART study won’t tell us *how* these activities slow the spread of HIV (or if these activities don’t work, why they didn’t work). That is the purpose of the study you are now being asked to join, called the phylogenetics study. I will now explain more about the phylogenetics study.

The virus in one person with HIV is slightly different from the virus in another person. This study will use special laboratory tests to compare the different types of HIV in people in communities where PopART is being done. The purposes of this testing are to:

- Learn if the study has made a difference in the way HIV spreads in the community.
- Give us a good idea of when people first became infected with HIV.
- Help us understand whether people with HIV got their infections from people inside the community or outside the community
- Help us understand whether people with HIV in the community got their infections from people who have had HIV for a long time, or from people who have only been infected recently.
- Provide information about whether some HIV in the community is resistant to the drugs used to treat HIV infection.

What will happen during this study?

If you agree to participate, you will have some additional tests done on the left over blood from the samples that are already being collected from you to measure how HIV is affecting your health (CD4 count).

We will also ask you a few questions about when you were first tested positive for HIV and if you have taken any medications to fight HIV before. If you are a man, we will ask you if you have been circumcised or not. If you are a woman, we will ask you if you are pregnant. We will also ask for your age, if you are married or have a regular partner, where you live, your gender, and if you are participating in the PC study.

The additional tests that will be done on your left over blood sample will include tests such as the amount of HIV in your blood and whether the HIV you have can be treated by most ARVs. Comparing different types of virus found in blood of different people in the community will help researchers understand how HIV spreads in the communities.

If you decide to take part in this study, we may ask you to update some of this information every time you come to this clinic for routine HIV care until the PopART study is over. This information will include whether or not you are using anti-HIV drugs and your CD4 count.

These study procedures (collection of leftover blood samples and short questionnaire) will take place each time you visit the clinic for routine care until the end of the PopART study (end of 2016). For most patients, this is every 6–12 months but may vary based on your health. You will not be asked to attend any additional visits outside of your normal health care visits.

If you do not agree to participate in the study, the blood samples you provide today will only be used for your routine health care and anything left over will be destroyed. It will have no effect on the regular medical care that is available at this clinic and your CD4 count will still be taken for routine care.

There is no cost to you or payment provided for participating in this study. Test results will not be returned because all of the data will be anonymous (we will remove all personal identifying information).

Who else will be asked to participate in this study?

If they agree to participate, we will ask other patients who receive HIV care but have not yet started ARVs at the health care facilities (clinics) to participate in this study at eight other communities in Zambia where we are doing the main study.

In addition, we will ask participants who are already a part of the Population Cohort of PopART to participate in this additional study. We will not collect extra blood from these people either. If you are a participant in the Population Cohort, you may have been or will be asked to participate in this study again since trial staff will not know that you have been to a local clinic.

We expect approximately 9,200 people to participate in this study.

All blood samples collected as part of this trial will be shipped and securely stored outside of the country for study-related testing. Samples will be stored indefinitely and used only for studies related to HIV as we discussed above.

Some blood samples will be shared with scientists working in other African countries, to help medical professionals learn about different HIV epidemics. In addition, some basic information such as age, HIV treatment status, and CD4 count will be provided.

What are the potential benefits?

There are no direct benefits to you, other than the satisfaction you may get from adding to the knowledge gained from this important study. This knowledge may help the research teams and health authorities worldwide understand how the PopART trial affected the spread of HIV in the study communities. It will help us to understand patterns of new HIV infections and how they spread through communities. It will also tell us how certain types of HIV medications help control the different types of HIV which people can have which can also help medical professionals make decisions about what kinds of HIV medicines to provide in your community.

What are the possible risks or discomforts?

There are no foreseen increased risks to being part of this additional study. You may feel anxious or embarrassed when sensitive questions are asked. You do not have to answer any question you do not wish to answer. There is an unlikely possibility that others may find out that you are participating in this study. In the unlikely event that this happens, you may experience stigma or other personal problems with family, friends, and people in your community. We will do everything we can to protect your privacy.

How will my confidentiality and privacy be protected?

The main concern that people may have is privacy. People may worry that a researcher would be able to know that one specific person infected another specific person with HIV. The study team will make sure that this is very unlikely to happen. If you are infected, we will make sure that no one is able to connect any information we learn about your HIV to you. Any information we have that would identify the blood samples as yours, like your name, will be permanently removed from the samples before tests are done on them for the additional study. None of the researchers or anyone in your community will ever be able to know that the samples of blood are linked to you. The only information that will link to your blood sample will be a study identification number which will be put on the tubes containing the blood, but it will not be possible to work out your identity from this information. In addition, we will never publish the name of your community in any results.

People who may review the study records include: [insert name of site IRB/EC], local regulatory agencies, US National Institute of Allergy and Infectious (NIAID) contractors, US National Institutes of Health (NIH), study staff, and study monitors. Institutional Review Boards (IRBs) or Ethics Committees (ECs) are committees that watch over the safety and rights of research participants. Provincial and/or national public health officials may be given community-wide results but not individual results.

What happens if participating in this study injures me?

It is very unlikely that you could be injured as a result of participating in this study. Nothing that we will be asking of you should place you at risk for injury. However, if you are injured while participating in this study, you will be given immediate treatment for your injuries. You will not have to pay for this treatment. There is no program for compensation through the United States National Institutes of Health. However, in case of any research related injury, the London School of Hygiene and Tropical Medicine would provide compensation

Are there any alternatives to participation?

If we become aware of other research studies happening in this clinic that may benefit you, we will tell you about them.

What are some reasons why I may be withdrawn from this activity without my consent?

You may be withdrawn from the study without your consent if this research study or the PopART trial is stopped or canceled

Persons to Contact for Problems or Questions

If you have any questions about your participation in this research study, your rights as a research participant, or if you feel that you have experienced a research-related injury, contact:

Site research staff: (site insert name of the investigator or other study staff)

Research site address (es): (site insert physical address of above)

Daytime telephone number(s): (site insert telephone number)

24-hour contact number(s): (site insert telephone number)

If you have any questions or concerns about your rights as a research participant or want to discuss a problem, get information or offer input, you may contact:

Independent Review Board/Ethics Committee:

(Site insert name or title of person on the IRB, EC or other organization appropriate for the site)

Address of Independent Review Board: (site insert physical address of above)

Daytime Telephone Number: (site insert telephone number of above)

Thank you for reading this information sheet. If you have any questions, please ask them now. The interviewer will be pleased to answer them. If you wish to take part, please read and sign the consent form. We will offer you a copy of this this information sheet and encourage you to please keep it in a safe place.

**INFORMED CONSENT FORM FOR HCF PATIENTS
PHYLOGENETIC ANCILLARY STUDY**

HPTN 071-2 Phylogenetics in HPTN 071: An ancillary study to “Population Effects of Antiretroviral Therapy to Reduce HIV Transmission (PopART): A cluster-randomized trial of the impact of a combination prevention package on population-level HIV incidence in Zambia and South Africa”

- I have been given sufficient time to consider whether to take part in this study.
- My taking part in this research study is voluntary. I may decide not to take part or to withdraw from the research study at any time without penalty or loss of benefits or treatment to which I am entitled.
- The research study may be stopped at any time without my consent.
- I have had an opportunity to ask the study staff questions about this research study. My questions so far have been answered to my satisfaction.
- I have been told that my participation in the additional study involves additional testing of my blood samples that will already be collected for measuring my CD4 count today at the clinic.
- I have been told who else will be invited to participate in this study, and that some blood samples will come from people who are in the main PopART trial.
- I have been told what the possible risks and benefits are from taking part in this research study. I will not benefit individually if I take part in this research study.
- I do not give up my legal rights by signing this form.
- I have been told that before any study related procedures begin, I will be asked to voluntarily sign this Participant Information and Consent Form.
- I will receive a signed and dated copy of this Participant Information and Consent Form.
- I understand that my samples may be sent out of the country for study related testing.

I voluntarily agree to take part in this research study.

Participant’s Name (print) Participant’s Signature or Thumbprint Date

I certify that the information provided was given in a language that was understandable to the participant.

Name of Study Staff Study Staff Signature Date
Conducting Consent (print)

Witness’ Name (print) Witness’ Signature and Date Date
(If necessary)

APPENDIX 2: POWER CALCULATIONS

We explore the statistical power of the phylogenetic study to estimate epidemiological parameters of interest. We focus here on estimating the proportion of transmissions occurring during acute or early HIV infection (AEHI) since this is one of the important aims of the study. Similar reasoning could be applied to other aims of the study. Power calculations are based on the projected number of samples shown in Table 4.

To perform the power calculations, we use a simple, conservative “transmission pair” method. Transmissions cannot be proven with phylogenetics, but it is possible to estimate which transmission events are more likely than others. The method proceeds by identifying, for each incident case, whether there are phylogenetically likely transmitters among the sampled population. We then estimate what proportion of these likely transmitters experienced AEHI during the time period when transmission is likely to have occurred. Individuals in AEHI are identified either as seroconverters in the PC or based on the multi-assay algorithm in the HCF.

Based on the predicted sampling coverage, we calculated the expected number of incident cases for whom we expect to have a sequence and for whom a sequence is also available for the actual transmitting individual. We call these sequenced transmission pairs. We hypothesize that 75% of all sequenced transmission pairs can be identified in practice, using the viral phylogeny inferred from full genome sequence data. We will not generally be able to determine the direction of transmission for pairs of individuals where both are in early infection near the time of sequencing, but regardless, we can infer that one transmission during early infection is likely to have occurred.

This method is conservative, because it only considers patterns among sequenced transmission pairs rather than phylogenetic clusters or full phylogenies among sequenced individuals.

The contribution of AEHI to net transmission is uncertain [35], with two studies in sub-Saharan Africa reporting 9% and 38.6% (18.6-52.3) of transmissions attributable to individuals in AEHI, respectively [37, 56]. For these estimates (approximately 10% and 40%), the expected number of sampled transmission pairs is provided in table A1. These proportions are expected to increase during the intervention as incidence decreases, since transmission during AEHI is less likely to be prevented by PopART interventions. We therefore considered a range of values for the total proportion of transmissions due to AEHI, ranging from 5% to 60% as input parameters in our simulations, and re-estimated these proportions based on imperfect detection. The best estimate and corresponding 95% binomial confidence intervals (calculated with the Wilson score method [58]) are shown in Figure A1.

Table A.1 Scenarios considered for estimating AEHI

Country	Arm	Target	Example input parameter – low %transmissions from AEHI (% AEHI)	Example input parameter – high % transmissions from AEHI (% AEHI)	Proportion of sex acts that occur with individuals outside the community. § (%)	Proportion of sequenced transmission pairs that can be identified with phylogenetic methods (%)	Expected number of sequenced transmission pairs that can be identified	Expected number of sequenced transmission pairs that can be identified and have an early transmitter in the lower baseline scenario	Expected number of sequenced transmission pairs that can be identified and have an early transmitter in the higher baseline scenario
ZA	A	central	10	40	13	75	83	8	33
ZA	B	central	10	40	7	75	149	14	59
ZA	C	central	10	40	5	75	37	3	14

§ Predicted with the best estimate PopART model under central and pessimistic targets for arm A and B.

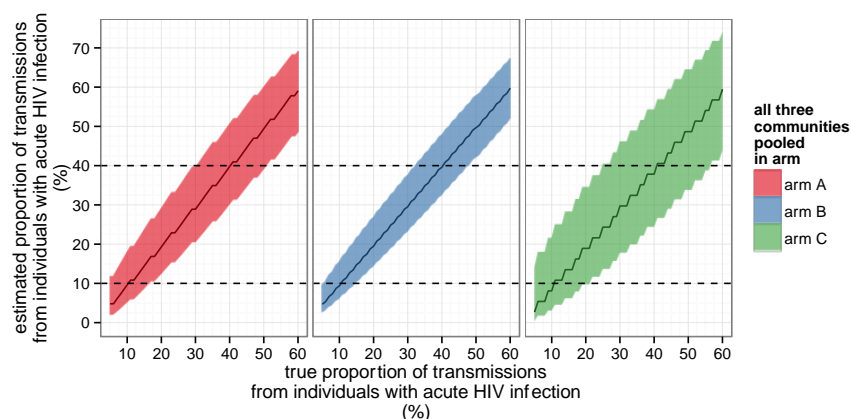


Figure A.1 Estimated % transmissions that occur during AEHI versus true % based on the ‘transmission pair’ method. The actual %AEHI values was varied from 5% to 60%, and re-estimated for each arm with the ‘transmission pair’ method under the assumptions in table 2 and A.1. Maximum-likelihood estimates (line) are shown along with 95% confidence intervals (colored regions).

Results

Table A.1 lists the predicted number of sequenced transmission pairs with a newly HIV infected individual that became infected during the trial by arm. Numbers are pooled across the three arm ‘A’, ‘B’ and ‘C’ communities in Zambia. Considering sequence data from an estimated total of ~9,400 individuals, these numbers are low and reflect the conservative estimation approach taken here.

Figure A.1 shows the estimated % transmissions that occur during AEHI, estimated with the ‘transmission pair’ method, against the true proportion that was used as input to the simulations. The 95% binomial confidence intervals are largest in arm C, followed by arm A, and smallest in arm B. To reject the baseline 10% AEHI scenario in arm C based on a 95% confidence interval, the true % should be approximately 20%, or equivalently the effect size should be approximately 10% or larger. As incidence is expected to decrease during the intervention, the proportion of transmissions due to AEHI is larger in arms A and B than in arm C. Considering the results in figure A.1, an effect size of approximately 7% is required to reject the baseline 10% AEHI scenario in arms A or B, and this effect size only increases to approximately 12% in order to reject the higher baseline scenario.

Summary

A high proportion of transmissions occurring during AEHI may negatively affect trial outcomes [39]. We found that the proposed design of the phylogenetic study can likely detect a 15% difference in the proportion of transmission during AEHI.

Statistical power was assessed with a conservative ‘transmission pair’ method. More sophisticated estimation methods [45, 50, 51] can be expected to improve substantially upon this method, and so the analysis provided should be regarded as a lower bound on the actual statistical power of the analysis. Such methods also explicitly model and adjust for issues of misclassification associated with missing data, for example in cases where two people were infected by a third person not in the sample, or when two individuals are part of a wider chain of transmission that is not observed. These types of misclassification are less of a problem when the epidemic predominantly features heterosexual transmission compared to analyses of epidemics amongst men who have sex with men, but nonetheless require careful consideration.

REFERENCES

1. Shankarappa R, Margolick JB, Gange SJ, Rodrigo AG, Upchurch D, Farzadegan H, Gupta P, Rinaldo CR, Learn GH, He X *et al*: Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J Virol* 1999, 73(12):10489-10502.
2. Alizon S, Fraser C: Within-host and between-host evolutionary rates across the HIV-1 genome. *Retrovirology* 2013, 10.
3. Rambaut A, Robertson DL, Pybus OG, Peeters M, Holmes EC: Human immunodeficiency virus. Phylogeny and the origin of HIV-1. *Nature* 2001, 410(6832):1047-1048.
4. Gray RR, Tatem AJ, Lamers S, Hou W, Laeyendecker O, Serwadda D, Sewankambo N, Gray RH, Wawer M, Quinn TC *et al*: Spatial phylodynamics of HIV-1 epidemic emergence in east Africa. *Aids* 2009, 23(14):F9-F17.
5. Veras NM, Santoro MM, Gray RR, Tatem AJ, Lo Presti A, Olearo F, Cappelli G, Colizzi V, Takou D, Torimiro J *et al*: Molecular epidemiology of HIV type 1 CRF02_AG in Cameroon and African patients living in Italy. *AIDS research and human retroviruses* 2011, 27(11):1173-1182.
6. Hue S, Hassan AS, Nabwera H, Sanders EJ, Pillay D, Berkley JA, Cane PA: HIV type 1 in a rural coastal town in Kenya shows multiple introductions with many subtypes and much recombination. *AIDS research and human retroviruses* 2012, 28(2):220-224.
7. Delatorre EO, Bello G: Phylodynamics of HIV-1 subtype C epidemic in east Africa. *PLoS One* 2012, 7(7):e41904.
8. Gilbert MT, Rambaut A, Wlasiuk G, Spira TJ, Pitchenik AE, Worobey M: The emergence of HIV/AIDS in the Americas and beyond. *Proc Natl Acad Sci U S A* 2007, 104(47):18566-18570.
9. Lewis F, Hughes GJ, Rambaut A, Pozniak A, Leigh Brown AJ: Episodic sexual transmission of HIV revealed by molecular phylodynamics. *PLoS Med* 2008, 5(3):e50.
10. Hughes GJ, Fearnhill E, Dunn D, Lycett SJ, Rambaut A, Leigh Brown AJ, Collaboration UHDR: Molecular phylodynamics of the heterosexual HIV epidemic in the United Kingdom. *PLoS Pathog* 2009, 5(9):e1000590.
11. Kouyos RD, von Wyl V, Yerly S, Boni J, Taffe P, Shah C, Burgisser P, Klimkait T, Weber R, Hirschel B *et al*: Molecular epidemiology reveals long-term changes in HIV type 1 subtype B transmission in Switzerland. *J Infect Dis* 2010, 201(10):1488-1497.
12. Bezemer D, van Sighem A, Lukashov VV, van der Hoek L, Back N, Schuurman R, Boucher CA, Claas EC, Boerlijst MC, Coutinho RA *et al*: Transmission networks of HIV-1 among men having sex with men in the Netherlands. *Aids* 2010, 24(2):271-282.
13. von Wyl V, Kouyos RD, Yerly S, Boni J, Shah C, Burgisser P, Klimkait T, Weber R, Hirschel B, Cavassini M *et al*: The role of migration and domestic transmission in the spread of HIV-1 non-B subtypes in Switzerland. *The Journal of infectious diseases* 2011, 204(7):1095-1103.
14. Ou CY, Ciesielski CA, Myers G, Bandea CI, Luo CC, Korber BT, Mullins JI, Schochetman G, Berkelman RL, Economou AN *et al*: Molecular epidemiology of HIV transmission in a dental practice. *Science* 1992, 256(5060):1165-1171.
15. Leitner T, Escanilla D, Franzen C, Uhlen M, Albert J: Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis. *Proc Natl Acad Sci U S A* 1996, 93(20):10864-10869.
16. Yirrell DL, Robertson P, Goldberg DJ, McMenamin J, Cameron S, Leigh Brown AJ: Molecular investigation into outbreak of HIV in a Scottish prison. *Bmj* 1997, 314(7092):1446-1450.
17. Metzker ML, Mindell DP, Liu XM, Ptak RG, Gibbs RA, Hillis DM: Molecular evidence of HIV-1 transmission in a criminal case. *Proc Natl Acad Sci U S A* 2002, 99(22):14292-14297.
18. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, Hakim JG, Kumwenda J, Grinsztejn B, Pilotto JH *et al*: Prevention of HIV-1 infection with early antiretroviral therapy. *N Engl J Med* 2011, 365(6):493-505.

19. Eshleman SH, Hudelson SE, Redd AD, Wang L, Debes R, Chen YQ, Martens CA, Ricklefs SM, Selig EJ, Porcella SF *et al*: Analysis of genetic linkage of HIV from couples enrolled in the HIV Prevention Trials Network 052 trial. *J Infect Dis* 2011, 204(12):1918-1926.
20. Campbell MS, Mullins JI, Hughes JP, Celum C, Wong KG, Raugi DN, Sorensen S, Stoddard JN, Zhao H, Deng W *et al*: Viral linkage in HIV-1 seroconverters and their partners in an HIV-1 prevention clinical trial. *PLoS One* 2011, 6(3):e16986.
21. Hue S, Pillay D, Clewley JP, Pybus OG: Genetic analysis reveals the complex structure of HIV-1 transmission within defined risk groups. *Proc Natl Acad Sci U S A* 2005, 102(12):4425-4429.
22. Brown AJ, Lobidel D, Wade CM, Rebus S, Phillips AN, Brettle RP, France AJ, Leen CS, McMenamin J, McMillan A *et al*: The molecular epidemiology of human immunodeficiency virus type 1 in six cities in Britain and Ireland. *Virology* 1997, 235(1):166-177.
23. Pao D, Fisher M, Hue S, Dean G, Murphy G, Cane PA, Sabin CA, Pillay D: Transmission of HIV-1 during primary infection: relationship to sexual risk and sexually transmitted infections. *AIDS* 2005, 19(1):85-90.
24. Oster AM, Pieniazek D, Zhang X, Switzer WM, Ziebell RA, Mena LA, Wei X, Johnson KL, Singh SK, Thomas PE *et al*: Demographic but not geographic insularity in HIV transmission among young black MSM. *Aids* 2011, 25(17):2157-2165.
25. Chen JH, Wong KH, Chan KC, To SW, Chen Z, Yam WC: Phylodynamics of HIV-1 subtype B among the men-having-sex-with-men (MSM) population in Hong Kong. *PLoS One* 2011, 6(9):e25286.
26. Levy I, Mor Z, Anis E, Maayan S, Leshem E, Pollack S, Chowers M, Mor O, Riesenberk K, Sthoeger Z *et al*: Men who have sex with men, risk behavior, and HIV infection: integrative analysis of clinical, epidemiological, and laboratory databases. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2011, 52(11):1363-1370.
27. Dennis AM, Hue S, Hurt CB, Napravnik S, Sebastian J, Pillay D, Eron JJ: Phylogenetic insights into regional HIV transmission. *Aids* 2012, 26(14):1813-1822.
28. Fisher M, Pao D, Brown AE, Sudarshi D, Gill ON, Cane P, Buckton AJ, Parry JV, Johnson AM, Sabin C *et al*: Determinants of HIV-1 transmission in men who have sex with men: a combined clinical, epidemiological and phylogenetic approach. *AIDS* 2010, 24(11):1739-1747.
29. Volz E, Ionides E, Romero-Severson E, Brandt MG, Mokotoff E, Koopman J: HIV-1 Transmission During Early Infection in Men Who Have Sex with Men: A Phylodynamic Analysis. *PLoS Med* 2013, in press.
30. Balode D, Skar H, Mild M, Kolupajeva T, Ferdats A, Rozentale B, Leitner T, Albert J: Phylogenetic analysis of the Latvian HIV-1 epidemic. *AIDS research and human retroviruses* 2012, 28(8):928-932.
31. Grabowski MK, Lessler J, Redd A, Kagaayi J, Laeyendecker O, Nalugoda F, Serwadda D, Wawer M, Quinn TC, Gray RH: S15.1 Frequent Viral Introductions Sustain Local HIV Epidemics in Rural Africa. *Sexually transmitted infections* 2013, 89(Suppl 1):A22-A22.
32. Redd AD, Mullis CE, Serwadda D, Kong X, Martens C, Ricklefs SM, Tobian AAR, Xiao C, Grabowski MK, Nalugoda F *et al*: The Rates of HIV Superinfection and Primary HIV Incidence in a General Population in Rakai, Uganda. *Journal of Infectious Diseases* 2012, 206(2):267-274.
33. Hollingsworth TD, Laeyendecker O, Shirreff G, Donnelly CA, Serwadda D, Wawer MJ, Kiwanuka N, Nalugoda F, Collinson-Streng A, Ssempijja V *et al*: HIV-1 transmitting couples have similar viral load set-points in Rakai, Uganda. *PLoS Pathog* 2010, 6(5):e1000876.
34. Pillay D, Rambaut A, Geretti AM, Brown AJ: HIV phylogenetics. *Bmj* 2007, 335(7618):460-461.
35. Cohen MS, Shaw GM, McMichael AJ, Haynes BF: Acute HIV-1 infection. *N Engl J Med* 2011, 364(20):1943-1954.
36. Ma ZM, Stone M, Piatak M, Jr., Schweighardt B, Haigwood NL, Montefiori D, Lifson JD, Busch MP, Miller CJ: High specific infectivity of plasma virus from the pre-ramp-up and ramp-up stages of acute simian immunodeficiency virus infection. *J Virol* 2009, 83(7):3288-3297.

37. Hollingsworth TD, Anderson RM, Fraser C: HIV-1 transmission, by stage of infection. *The Journal of infectious diseases* 2008, 198(5):687-693.
38. Quinn TC, Wawer MJ, Sewankambo N, Serwadda D, Li C, Wabwire-Mangen F, Meehan MO, Lutalo T, Gray RH: Viral load and heterosexual transmission of human immunodeficiency virus type 1. Rakai Project Study Group. *N Engl J Med* 2000, 342(13):921-929.
39. Cohen MS, Dye C, Fraser C, Miller WC, Powers KA, Williams BG: HIV treatment as prevention: debate and commentary--will early infection compromise treatment-as-prevention strategies? *PLoS Med* 2012, 9(7):e1001232.
40. Brenner BG, Roger M, Routy JP, Moisi D, Ntemgwa M, Matte C, Baril JG, Thomas R, Rouleau D, Bruneau J *et al*: High rates of forward transmission events after acute/early HIV-1 infection. *The Journal of infectious diseases* 2007, 195(7):951-959.
41. Drummond AJ, Suchard MA, Xie D, Rambaut A: Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 2012, 29(8):1969-1973.
42. de Oliveira T, Shafer RW, Seebregts C: Public database for HIV drug resistance in southern Africa. *Nature* 2010, 464(7289):673.
43. Gall A, Ferns B, Morris C, Watson S, Cotten M, Robinson M, Berry N, Pillay D, Kellam P: Universal amplification, next-generation sequencing, and assembly of HIV-1 genomes. *Journal of clinical microbiology* 2012, 50(12):3838-3844.
44. Henn MR, Boutwell CL, Charlebois P, Lennon NJ, Power KA, Macalalad AR, Berlin AM, Malboeuf CM, Ryan EM, Gnerre S *et al*: Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection. *PLoS Pathog* 2012, 8(3):e1002529.
45. Stadler T, Bonhoeffer S: Uncovering epidemiological dynamics in heterogeneous host populations using phylogenetic methods. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* 2013, 368(1614):20120198.
46. Volz E: Identifying HIV-1 transmission risk factors and the source of transmission from molecular data. In: *20th HIV Dynamics & Evolution Meeting: 2013; Utrecht, The Netherlands*.
47. van Zyl GU, van der Merwe L, Claassen M, Zeier M, Preiser W: Antiretroviral resistance patterns and factors associated with resistance in adult patients failing NNRTI-based regimens in the Western Cape, South Africa. *Journal of medical virology* 2011, 83(10):1764-1769.
48. van Zyl GU, Claassen M, Engelbrecht S, Laten JD, Cotton MF, Theron GB, Preiser W: Zidovudine with nevirapine for the prevention of HIV mother-to-child transmission reduces nevirapine resistance in mothers from the Western Cape, South Africa. *Journal of medical virology* 2008, 80(6):942-946.
49. Leigh Brown AJ, Lycett SJ, Weinert L, Hughes GJ, Fearnhill E, Dunn DT: Transmission network parameters estimated from HIV sequences for a nationwide epidemic. *The Journal of infectious diseases* 2011, 204(9):1463-1469.
50. Volz EM, Kosakovsky Pond SL, Ward MJ, Leigh Brown AJ, Frost SD: Phylodynamics of infectious disease epidemics. *Genetics* 2009, 183(4):1421-1430.
51. Rasmussen DA, Ratmann O, Koelle K: Inference for nonlinear epidemiological models using genealogies and time series. *PLoS Comput Biol* 2011, 7(8):e1002136.
52. Ratmann O, Donker G, Meijer A, Fraser C, Koelle K: Phylodynamic inference and model assessment with approximate bayesian computation: influenza as a case study. *PLoS Comput Biol* 2012, 8(12):e1002835.
53. Stadler T, Kuhnert D, Bonhoeffer S, Drummond AJ: Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proc Natl Acad Sci U S A* 2013, 110(1):228-233.
54. Drummond AJ, Suchard MA, Xie D, Rambaut A: Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol*, 29(8):1969-1973.
55. de Silva E, Ferguson NM, Fraser C: Inferring pandemic growth rates from sequence data. *Journal of the Royal Society, Interface / the Royal Society* 2012, 9(73):1797-1808.

56. Powers KA, Ghani AC, Miller WC, Hoffman IF, Pettifor AE, Kamanga G, Martinson FE, Cohen MS: The role of acute and early HIV infection in the spread of HIV and implications for transmission prevention strategies in Lilongwe, Malawi: a modelling study. *Lancet* 2011, 378(9787):256-268.
57. Van Zyl GU, Liu TF, Claassen M, Engelbrecht S, de Oliveira T, Preiser W, Wood NT, Travers S, Shafer RW: Trends in Genotypic HIV-1 Antiretroviral Resistance between 2006 and 2012 in South African Patients Receiving First- and Second-Line Antiretroviral Treatment Regimens. *PLoS One* 2013, 8(6):e67188.
58. Coull AAaBA: Approximate is better than "exact" for interval estimation of binomial proportions. *American Statistician* 1998(52):119–126.