How can we develop acceptable and effective phylogenetically informed approaches to HIV prevention in the UK?

Dr Larissa Mulka

This dissertation is submitted for the degree of Doctor of Philosophy, Brighton and Sussex Medical School

ABSTRACT

Due to technical and computational advances, the use of phylogenetics in HIV is evolving. Previously, analyses were time consuming and expensive allowing only retrospective analyses on relatively small scales, but now phylogenetic analyses can be, and are, used in near-real time to describe transmission dynamics and guide public health response, aiming to prevent further transmission. While phylogenetically underpinned prevention is being implemented in the USA as part of its epidemic response, there is less experience in the UK, and there has been little research into the ethical and acceptability issues in this context.

This thesis aims to identify how the use of real-time phylogenetics to guide HIV prevention interventions could be acceptably implemented in the UK. A literature review and two empirical studies were conducted to address two research questions: 1. How could phylogenetics be implemented in a way that is acceptable to patients?; 2. Can real-time phylogenetically guided interventions be piloted on a local scale, providing evidence for effectiveness and real-life acceptability?

A systematic review was conducted to identify and draw lessons from the use of approaches for case finding of infectious disease guided by phylogenetics, and the barriers, facilitators and ethical issues associated with this. A phylogenetic analysis sought to identify sources of HIV infection in incident cases of HIV in Brighton, to identify whether this was a relatively 'closed' cohort, and inform piloting of a real-time phylogenetically led prevention system. A qualitative study explored acceptability of the use of phylogenetics with key stakeholders, and potential negative outcomes of its use.

The findings of these studies were synthesised to inform intervention development. There has been a paucity of previous research or reporting of barriers, facilitators, and acceptability and ethical issues in the context of phylogenetic data used for purposes of infectious disease case finding. The acceptability of such data use to patients and healthcare staff depends on the perceived risks balanced against potential public health benefits of this use of data, and use of phylogenetics in a manner deemed unacceptable risks disengaging patients from testing for HIV. Acceptability of the use of phylogenetics to guide public health interventions increases as the potential prevention benefits increase, and is enhanced by certain protections, including strict security measures and limiting access to data, informed consent, protection from police use and improved understanding of the purposes and limitations of phylogenetic

analysis. Sources of transmission to patients attending Brighton's HIV centre were found to be geographically dispersed throughout the UK, suggesting piloting a phylogenetic system locally would limit its utility in terms of prevention benefits, and to be effective, such a system would require large scale, national input.

To introduce phylogenetically guided public health interventions in the UK, evidence of clear public health benefit is required, which may be achieved through nationwide anonymous real-time phylogenetic surveillance to identify whether clusters that may be disrupted by interventions exist, and will be supplemented by outcomes reported in the USA. If we gain evidence of benefit, strategies to mitigate risk require development, including protections from judicial use, approaches for informed consent and stringent security and data management policies. Introducing phylogenetics to guide public health interventions in the UK without addressing these considerations risks disengaging individuals from testing and on-going care.

Table of Contents

List of tables

List of Figures

Acknowledgements

I would like to sincerely thank my supervisors, Jackie Cassell, Jaime Vera and Andrew Leigh Brown for their constant support, encouragement and advice throughout my PhD. Thank you to Brighton and Sussex University Hospitals NHS Trust, who funded my PHD via the Clinical Investigations and Research unit. In addition I would like to thank the following people without whom this project would not be possible; Manon Ragonnet-Cronin and Emma Hodcroft for their phylogenetics advice and help with all of my coding errors; Alex Pollard, Shanu Sadwani and Duncan Shrewsbury for their advice and insightful comments on my qualitative work; Bobbie Farsides for her input and advice on this project; Edwin Bernard, Matthew Williams and Robert James for their legal expertise and for providing a community perspective on my project on the whole; the participants that gave their time to take part in my studies; and to my Brighton clinical colleagues at the Lawson Unit for providing data, assistance with recruitment and moral support. Lastly I would like to thank Martin Fisher for his inspiration and encouragement in applying for this position.

I would like to thank my father, Orest, for proofreading my writing at very short notice on multiple occasions and my wonderful friends and family for their continual support.

List of abstracts from this thesis

1)**Mulka L**, Ragonnet-Cronin M, Tostevin A, Dunn D, Leigh-Brown A on Behalf of the UK HIV Drug resistance Database. Identifying the sources of transmission in a UK HIV-positive cohort. 24th International HIV Dynamics and Evolution. Isle of Skye, May 2017

2) **Mulka L**, Vera J, Leigh-Brown A, Cassell J. How can we use phylogenetics to facilitate clinical case finding and partner notification in HIV? Lessons from a systematic review of its use in stigmatized infectious diseases. ISSTDR, Rio de Janeiro, July 2017

3) **Mulka L**, Ragonnet-Cronin M, Vera JH, Cassell J, Tostevin A, Dunn D, Leigh Brown A on behalf of the UK HIV Drug Resistance Database. Are Real-Time Phylogeny Guided Interventions Feasible? A Longitudinal Analysis. Conference on Retroviruses and Opportunistic Infections. Boston, Massachusetts, March 2018

4)Northam A, **Mulka L**, Cassell J. Ethical Tensions in the use of phylogenetic analysis of HIV transmission networks: a scoping review. Fourth Joint Conference of BHIVA and BASHH. Edinburgh, April 2018

5)**Mulka L,** Pollard A, Vera JH, Leigh Brown A, Farsides B, Bernard E, Williams, M, Cassell J. What is the acceptability of using phylogenetic data in clinical and public health practice? BHIVA Spring conference, Bournemouth, 2019

Abbreviations and Acronyms

AIDS	Acquired immune deficiency syndrome
ART	Antiretroviral therapy
BEAST	Bayesian evolutionary analysis by sampling trees
BHIVA	British HIV Association
BLAST	Basic local alignment tool
CDC	Centres for disease control (USA)
CHIPS	Collaborative HIV paediatrics study
CIOMS	Council for international organizations of medical sciences
CRF	Circulating recombinant form
DNA	Deoxyribonucleic acid
DRM	Drug resistance mutation
DSPS-HIV	Discrete spatial phylo simulator
ESS	Effective sample size
ETD	Estimated transmission date
FGD	Focus group discussion
GDPR	General data protection regulation
HAART	Highly active antiretroviral therapy
HAND	HIV & AIDS new diagnoses database
HARS	HIV and AIDS reporting system
HIV	Human immunodeficiency virus
IDI	In-depth interview
INI	Integrase inhibitor
LANL	Los Alamos National Laboratories
LTB	Latent tuberculosis
MCC	Maximum clade credibility
MRCA	Most recent common ancestor
MSM	Men who have sex with men
MTCT	Mother to child transmission
NGS	Next generation sequencing
NJ	Neighbour Joining
NNRTI	Non-nucleotide reverse transcriptase inhibitor
NRTI	Nucleotide reverse transcriptase inhibitor
PEP	Post-exposure prophylaxis
PHE	Public Health England
PHI	Public health interventions
PI	Protease inhibitor
PLWH	People living with HIV
PN	Partner notification
PrEP	Pre-exposure prophylaxis
PWID	People who inject drugs
RHI	Recent HIV infection

RITA	Recent infection testing algorithm
RNA	Ribonucleic acid
SIV	Simian Immunodeficiency virus
SOPHID	Survey of prevalent HIV infections diagnosed
STI	Sexually transmitted infection
TaSP	Treatment as prevention
TB	Tuberculosis
TDR	Transmitted drug resistance
TNS	Transmission network score
UK CHIC	UK collaborative HIV cohort
UK RDB	UK HIV drug resistance database
UNAIDS	Joint United Nations programme on HIV/AIDS
UPGMA	Unweighted pair group method with arithmetic means
VL	Viral load
WHO	World health organisation
WPGMA	Weighted pair group method with arithmetic mean

Declaration

I, Larissa Mulka, declare that the research contained in this thesis, unless otherwise formally indicated within the text, is the original work of the author. The thesis has not been previously submitted to these or any other university for a degree, and does not incorporate any material already submitted for a degree.

Signed

Dated 8th December 2019

Chapter 1. Introduction

Chapter purpose and summary

The aim of this introduction is to provide a general overview of the subjects this thesis focuses on, the HIV epidemic, transmission and prevention, and the use of phylogenetics, to provide context for the subsequent chapters. This chapter provides an overview of HIV as an epidemic; describing the evolution of the epidemic, providing an overview of the current UK epidemic and identifying factors associated with transmission and prevention in order to clarify actions required to achieve control and highlight the challenges in accomplishing this. Furthermore, it provides an explanation of phylogenetics, particularly in the context of HIV 'transmission networks', provides an overview of the processes required to perform analyses and describes opportunities for phylogenetics in HIV prevention.

1.1 The HIV epidemic

Human Immunodeficiency virus (HIV) is retrovirus affecting humans that if left untreated causes progressive failure of the immune system by attacking a type of lymphocyte, the CD4 T-cell, leading to acquired immunodeficiency syndrome (AIDS), characterised by the manifestation of AIDS defining opportunistic infections or malignancies.

AIDS was first recognised in 1981, after a cluster of homosexual males presenting with unusual conditions associated with immunodeficiency were identified in New York City and California, connected by sexual contact [1, 2, 3] and the hypothesis of a sexually transmitted immunodeficiency was proposed. HIV is now a global pandemic, and has caused 32 million deaths worldwide[4], however improvements in understanding of the virus, and rapid advances in treatment and care mean HIV has turned from a mostly fatal disease to a treatable and preventable condition with excellent outcomes.

1.1.1 The global burden of HIV

There were an estimated 37.9 million people living with HIV (PLWH) globally at the end of 2018, of which the majority were found in sub-Saharan Africa, an epidemic driven mainly by heterosexual sex[4]. Overall numbers of new infections and AIDS related deaths have been gradually decreasing with the increased uptake of antiretroviral therapy (ART), however HIV incidence in Eastern Europe and central Asia has increased by 29% since 2010, largely driven by injecting drug use[5].

In 2016, UNAIDS adopted a new strategy to end the AIDS epidemic as a public health threat by 2030[6]. This is focussed around 8 key areas; promoting health and wellbeing with access to HIV testing and treatment for all children, adolescents and adults living with HIV; eliminate mother to child transmission (MTCT) and ensure mothers' health and wellbeing is sustained; reduce inequalities for young people, allowing them access to services and empowering them to protect themselves against HIV; ensuring tailored HIV combination prevention services to key populations including sex-workers, men who have sex with men (MSM), people who inject drugs (PWID) prisoners, transgender people and migrants; promoting gender equality to end gender based and intimate partner violence; removal of laws, policies and stigma that block effective responses to HIV; ensuring adequate funding and implementation of AIDS responses; and ensuring the integration of people-centred HIV and health services. Within this framework, UNAIDS aims for fewer than 500,000 new HIV infections, fewer than 500,000 AIDS related deaths and elimination of HIV-related discrimination by 2020. Preliminary data shows progress in achieving 90-90-90 targets (90% PLWH diagnosed, 90% of diagnosed on treatment, 90% on treatment to have an undetectable viral load[7]), with 79% of PLWH globally knowing their status, 78% of whom are accessing treatment with 86% having a supressed viral load in 2018. However, the reduction in incidence remains far behind the target of fewer than 500,000 new infections in 2020, with approximately 1.7 million new infections in 2018[5].

1.1.2 Epidemiology of HIV in the UK

In 2017 for the first time, the UK met the UNAIDS 90:90:90 targets[8]. Of the 101,600 people living with HIV infection in the UK, 92% were diagnosed, 98% of whom were on

treatment, with 97% of those on treatment being virally supressed.

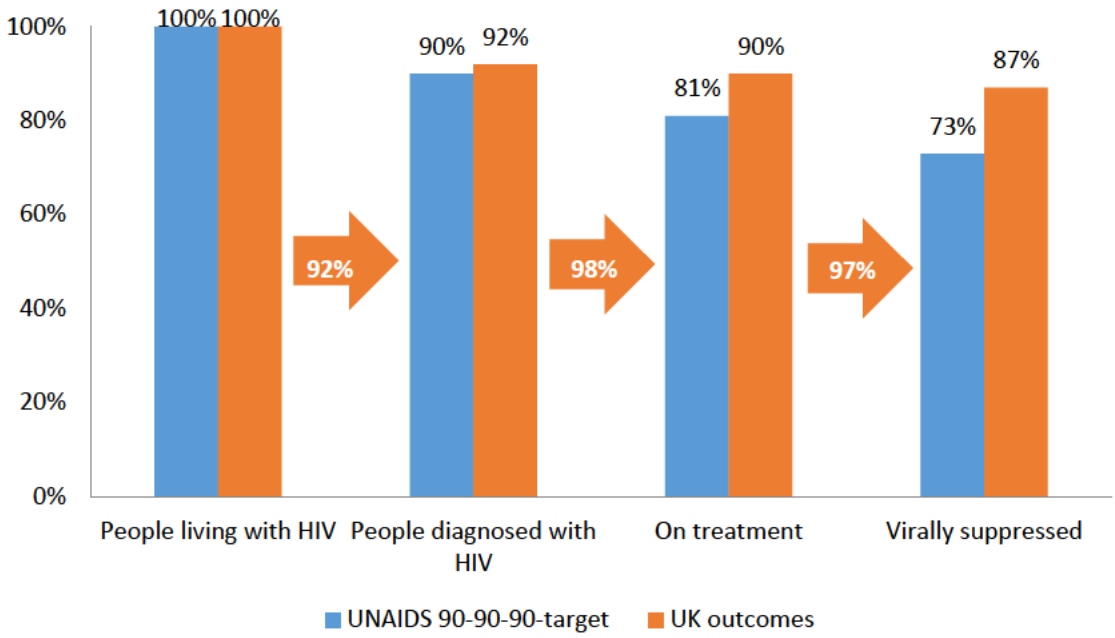


Figure 1: **Continuum of HIV care in the UK in relation to UNAIDS 90-90-90 targets, 2017. Adapted from Nash et al, 2018[8].**

The largest risk groups living with HIV in the UK are gay and bisexual men (48,900) and heterosexual men and women (47,400) of which black African men and women make up the largest proportion (57%). Approximately 2,500 people who inject drugs live with HIV in the UK. People living with HIV in the UK are getting older, largely due to the success of ART, with the median age increasing from 40 to 46 from 2008 to 2017, with almost 40% of patients receiving care being age 50 or above.

1.1.2.1 New infections

There was a decline in HIV incidence in both MSM and heterosexuals in 2017. This has been particularly apparent amongst MSM. The incidence in this group has almost halved since 2012, decreasing by almost a third since 2015, most strikingly in the 'London large fall clinics'[1], where there was a reduction of 46% in two years. Although new diagnoses amongst black African and black Caribbean heterosexuals have been steadily decreasing

[1] Dean Street, Mortimer Market, Homerton, St Mary's, Guys' and St Thomas'

over the past decade, 2017 saw the first declines amongst non-black African or Caribbean heterosexuals, with a particularly significant reduction in white heterosexual men. This may be in part due to the on-going decreases in MSM, with evidence that some self reported heterosexual men may have acquired their infection from men in phylogenetic analyses[9]. In addition, there was a decline in the number of late diagnoses from 3,895 in 2008, to 1,879 in 2017, although 43% of diagnoses in 2017 were still made at a late stage (CD4 T cell count < 350 cells/mm^3). Most new diagnoses in that year were amongst MSM (53%), and the largest proportion were in London (36%)[8].

1.1.2.2. Combination prevention

The recent declines in transmission are particularly noticeable given the lack of decline following the earlier widespread adoption of effective ART, and have been attributed to the implementation of *combination* HIV prevention; a combined approach of behavioural, biomedical and structural strategies in the context of well researched local epidemiology[10]. Strategies include condom provision, pre-exposure prophylaxis (PrEP) (section 1.3.5), frequent testing and early ART, and may be tailored to an individual to achieve maximal impact in terms of risk reduction (see section 1.3).

1.1.2.3 Undiagnosed infection

Despite progress with reductions in incidence, an estimated 7,800 (CI 5,600 to 12,600) individuals still remained undiagnosed in 2017 (8% of the population living with HIV). Approximately 5% of PLWH in London were undiagnosed, though 11% elsewhere in the UK had undiagnosed infection, with over half of the undiagnosed population being MSM. This means on-going prevention efforts are still required to further reduce this number, a group at risk of on-going transmission and late diagnosis, if we aim to end the epidemic in the UK.

1.2 Transmission of HIV; modes and associated factors

1.2.1. Modes of transmission

The most common route of HIV-1 transmission is transmucosal, usually via sexual contact, however the relative importance of each mode of transmission varies between populations. The commonest modes of transmission are sex between men, sex between men and women, intravenous drug use and mother to child transmission, though may be possible through transmucosal splashes to the eye[11]. Transmission by blood transfusion, infusion of infected untreated Factor VIII and IX, and iatrogenically through infected equipment has occurred in the past, though has been largely eliminated due to blood borne virus screening of donors, and infection control measures. Case studies exist where HIV has been transmitted via human bites, however this has occurred in the context of advanced infection with a high viral load, and evidence of deep wounds and blood in the oral cavity[12] and remains an unlikely transmission route. When dry, HIV becomes non-viable within hours, and has been shown to be non-viable within discarded syringes within 24 hours[13].

1.2.1.1 Sex between men

Sex between men accounts for the majority of new HIV infections in the UK. In addition to increased high-risk behaviour in this group, including multiple and casual partnerships, condomless sex and sex under the influence of drugs (Chemsex), sexual practices more prevalent in MSM such as anal intercourse increase risk of transmission (1:90 risk of transmission for receptive anal sex vs. 1:1000 for receptive vaginal sex with a person with untreated HIV[14]). Owing to the delicate nature of the rectal mucosa, anal penetration, with a penis, fist or sex toy carries a higher risk of trauma and bleeding, increasing the risk of transmission, and promotion of safer sexual practices within this group remains a priority health intervention in the UK.

1.2.1.2 Sex between men and women

Heterosexual sex remains the commonest mode of transmission worldwide, particularly amongst black African men and women, and accounts for approximately half of infections in the UK, though the incidence amongst all groups is decreasing[8].

1.2.1.3 Intravenous Drug use

Though relatively uncommon in the UK owing to needle exchanges and other harm reduction methods, sharing injecting equipment has been associated with localised outbreaks in high income countries[15, 16]. This was also the mode of spread within a notable recent outbreak in Glasgow[17] involving over 100 individuals, and is the source of major epidemics in Eastern Europe and Asia[5].

1.2.1.4 Mother to child transmission

MTCT of HIV may occur during pregnancy, childbirth or breastfeeding. Without treatment, overall transmission rate is 25%[18], however MTCT is largely preventable with appropriate interventions. These include the use of antiretroviral therapy, viral load monitoring, caesarean section if required, post exposure prophylaxis in the neonate and avoidance of breastfeeding, and with the introduction of opt-out HIV testing for all pregnant women in the UK, the risk of mother to child transmission in the UK is now less than 0.27%[19], with no MTCT to children born in the UK in 2017[8]. MTCT remains a significant challenge in resource poor countries. The UNAIDS *Global Plan towards the elimination of new HIV infections among children and keeping their mothers alive* however, has led to a reduction in transmission within 21 countries in sub-Saharan Africa from 28% to 14% by 2015 through increased access to antiretroviral therapy however [20], and its *Start Free, Stay Free, AIDS Free* framework continues to work to reduce this fraction further[21].

1.2.2. Factors influencing transmission

Several factors, both behavioural and biological, have been found to influence the likelihood of transmission of HIV, and are potential targets for prevention interventions.

1.2.2.1 Behavioural and condom use

Modifying sexual behaviours plays a crucial role in the prevention of new HIV infections [22]. Higher numbers of sexual partners and increased frequency of sexual acts, alongside type of sexual contact (particularly unprotected anal intercourse) are associated with a higher risk of HIV transmission [23]. Condoms, when used effectively are highly protective against HIV transmission[24, 25] and have likely significantly limited its spread during the epidemic. Chemsex, the use of recreational drugs during sexual activity has been documented in up to 16.5% of MSM attending sexual health clinics in London, and has been associated with increased risk behaviours and HIV diagnoses[26]. Advice and appropriate referral to drug services is therefore crucial within this group.

1.2.2.2 High viral load in the source

Increase in viral load (VL, the amount of virus in the blood) has been shown to directly correspond with transmissibility of HIV infection, with good evidence that having an undetectable viral load means the virus cannot be passed on through unprotected sexual intercourse (known as U=U, undetectable equals untransmissable)[27, 28, 29, 30, 31, 32]. Strategies to reduce viral load, including effective early treatment will have an impact on transmissibility, though the contribution of viral load to transmission remains a major issue amongst those who are undiagnosed, which, amongst MSM in the UK account for the majority of the 'infectious population'[33], reinforcing the need for frequent testing and condom use for those at risk.

1.2.2.3 Sexually transmitted infections

Alongside being a marker of higher risk sexual behaviour, concomitant sexually transmitted infections (STI) have been shown to heighten the risk of both transmitting and acquiring HIV infection[23]. Infectiousness may be via increased viral shedding[34, 35] and effects on HIV viral replication[36]. Susceptibility may be increased through proliferation of HIV target cells such as CD4 in the presence of an STI[37] or via breaches in the mucosal surface, for example with Herpes Simplex infection. Although not strictly an STI, bacterial vaginosis, a condition characterised by loss of normal lactobacilli and

overgrowth of commensal anaerobic bacteria in the vagina, has also been shown to increase susceptibility through disturbance in the protective vaginal flora[38]. Any diagnosis of an STI should prompt HIV testing[39], and consideration of retesting after the window period for HIV should be considered in higher risk cases (for example rectal STI in MSM).

1.2.2.4 Circumcision

Circumcision has been shown to reduce HIV acquisition in heterosexual men in high prevalence countries[40, 41]. Although a meta-analysis of observational studies has suggested little benefit in MSM[42], a subsequent cohort study has shown a significant reduction in HIV acquisition per episode of unprotected insertive anal intercourse in circumcised MSM in Australia[43]. This effect is thought to be due to the abundance of HIV target cells, including CD4 cells, Langerhans cells and macrophages located on the delicate inner mucosal surface of the foreskin[44, 45].

1.3 Prevention of HIV transmission

1.3.1 Testing

The British HIV Association (BHIVA) recommends HIV testing for all attendees of sexual health clinics, all patients presenting with indicator conditions, and all patients presenting to general practice, A&E and all hospital admissions in high prevalence areas, alongside screening in specialist services including antenatal, hepatitis and tuberculosis (TB) clinics. It is recommended that all MSM test yearly, or every three months if having unprotected sex with new or casual partners. Frequent testing means diagnoses are made, and effective treatment initiated earlier, reducing the likelihood of onward transmission[46], as well as providing opportunity for risk reduction counselling and condom provision. Testing activity continues to increase throughout the UK[8] and is likely to have contributed to the recent reductions in incidence seen, particularly amongst MSM, although 77% of new diagnoses in the UK in 2017 had not tested in the preceding 2 years[8].

1.3.2 Behavioural

Counselling and education to promote condom use and reduce risk behaviours are key for both primary and secondary prevention of HIV infection. Education amongst risk groups in the community, schools and within sexual health clinics provide potential opportunity to reduce the prevalence of unprotected sex, multiple partnerships and other high risk behaviours such as ChemSex. Behavioural interventions have been shown to increase condom use[47, 48], though randomised controlled trials of behavioural interventions have failed to show significant reductions in HIV incidence[49]. Nevertheless these remain an important component of HIV prevention.

1.3.3 Antiretroviral therapy (Section 1.5)

Lower viral loads reduce the risk of HIV transmission. Although irrefutable evidence for U=U has only been recently published, studies have demonstrated for some time that the risk of transmitting HIV whilst on suppressive ART is negligible, if possible at all[31], and treatment as prevention (TaSP) has been suggested in serodiscordant couples within the British HIV Association (BHIVA) guidelines since 2008[50]. Since the introduction of immediate ART to all patients following the release of the START (Strategic Timing of Antiretroviral Therapy) study results, showing significant benefit in morbidity and mortality with immediate initiation of ART irrespective of CD4 count[51] combined with the widespread publicising of U=U, the rates of suppression on ART in the UK are increasing, and the onward transmission of HIV should in theory continue to reduce.

1.3.4 Post-exposure prophylaxis

Post exposure prophylaxis (PEP) is the provision of ART to HIV negative individuals following a significant HIV exposure. The recommended regimen is combined tenofovir disoproxil/emtricitabine and raltegravir, an integrase inhibitor, for 28 days, initiated within 72 hours of exposure[14]. Although no randomised controlled trials of PEP have been conducted, animal studies[52], case-control studies of occupational PEP[53], and observational studies of PEP in neonates born to HIV positive mothers[54] have demonstrated efficacy in preventing transmission.

1.3.5 Pre-exposure prophylaxis

Pre-exposure prophylaxis (PrEP) has been available in the UK through the clinical trial PROUD[55] since 2015, and via the internet since 2016. Currently in the UK PrEP is available in the form of an oral single tablet regimen of tenofovir disoproxil/emtricitabine (commonly the brands Truvada, or Mylan) though other formulations are in development and evaluation. In tablet form, this is taken once daily for HIV-negative, at-risk individuals, or as an on-demand regimen for MSM. The PROUD trial, a study of once daily Truvada demonstrated an 86% reduction in HIV acquisition, as did the IPERGAY trial in France and Canada, investigating 'on-demand' Truvada[56]. PrEP is not currently available through the NHS, though NHS England has funded IMPACT, a major trial to assess PrEP eligibility, uptake and duration of use among Genitourinary Medicine attendees, currently recruiting its second round of participants due to high demand throughout the UK (www.prepimpacttrial.org.uk). Currently, sexual health services also support those privately purchasing PrEP, usually via the internet through websites such as https://www.iwantprepnow.co.uk. Numbers of MSM taking PrEP quadrupled in the UK in 2016, in line with its online availability, reaching an estimated 3,000 by the end of the year[57], a figure that continues to increase dramatically with recruitment for the IMPACT trial, and increasing awareness in those at risk.

1.3.6 Partner notification

Partner notification (PN), the process of identifying and offering testing to potentially exposed contacts of HIV is recommended for all newly diagnosed patients[58]. The main aims of PN are earlier identification of infected contacts, leading to earlier treatment/intervention and reduced onward transmission, and to increase awareness, and encourage prevention strategies in contacts that are found to be negative[59]. When performed effectively partner notification is successful in identifying new diagnoses of HIV, with a HIV prevalence of 20.9% amongst tested contacts in a recent UK national audit[60]. However, regional outcomes are variable with a prevalence of HIV of 10-37% in contacts in local audits[61].

Partner notification is often a difficult process for patients. Concerns may include loss of

confidentiality regarding their diagnosis, fears of stigma and blame from those contacted and guilt surrounding potential transmission to others, which may lead to non-disclosure of potentially infected contacts, particularly if non-regular. The rates of partner notification for ex- or casual partners have been shown to be much lower than in regular partners, and though the HIV prevalence in these former contacts has also been found to be lower, evidence suggests those with casual partners have greater numbers of casual partners than those with regular partners meaning a larger network of potentially untraceable infected contacts[62]. This is idea is echoed by a study reviewing partner notification in Chlamydia, which indicated the number of partners needing treatment to interrupt a transmission chain was fewer for casual partners compared with regular partners (1.1 vs. 2.5), though it is acknowledged this process is more resource intensive[63]. Earlier diagnosis also appears to have influence on the success of partner notification, with patients presenting with recent HIV able to name proportionately more contacts than those with established infection, with a higher rate of newly diagnosed contacts identified though partner notification in acute infection[64].

Untraceable contacts remain a significant barrier to PN, with approximately 12% of contacts to newly diagnosed PLWH being untraceable in the UK[60]. With the advances of the internet, and mobile phone dating 'apps' such as Gaydar and Grindr, online methods have complemented more traditional options for sexual networking, with almost half of MSM in the USA and UK having initiated sexual encounters online even back in 2005[65]. Internet based partner selection has been found to be associated with increased likelihood of unprotected anal intercourse, with serosorting (intercourse with partners thought to share the same HIV status) and strategic positioning with serodiscordant partners (adopting sexual positions with a lower risk of acquisition)[66], increasing the risk of HIV transmission. Both sex-on-premises venues and internet initiated sexual encounters may increase the likelihood of anonymous, uncontactable partners, posing further difficulties for partner notification.

1.3.7 Misconceptions and stigmatisation of HIV in the UK

Stigma has remained an issue in HIV since it was labelled a 'gay disease' in the early 1980s (Gay Related Immune Deficiency, GRID)[67]. Its association with drug use and sexual activity has led people in the past to believe brought on by individuals' own behaviours, and is therefore deserved[68].
Although the mass campaigns, including the UK governments *Don't' Die of Ignorance* campaign, featuring a tombstone bearing the word AIDS are likely to have resulted in a increase in testing at the time[69], fear based messages may promote stigma amongst both those living with HIV, and those belonging to risk groups[70], causing further marginalisation and acting as a barrier to testing and effective care.

Although stigma appears to be reducing with the introduction of effective and acceptable treatment, particularly amongst those at risk of HIV, ignorance relating to the prognosis of HIV, routes of transmission and U=U still remain. A recent survey conducted by the Terrence Higgins Trust and YouGov revealed lack of knowledge and belief in 'HIV myths' are still prevalent in the UK general public, with 20% still believing HIV can be transmitted through kissing[71]. Owing to a lack of public campaigns since those in the 1980s, ignorance also remains regarding the extent of the epidemic, with 53% of the public disagreeing that there is currently an epidemic in the UK, and worryingly, 68% of MSM share this view[71]. Such misconceptions prevent 'normalisation' of living with HIV, perpetuating stigma and isolation of PLWH.

HIV related stigma may lead to reluctance to test[72, 73], to disclose HIV status to friends or partners[74], and to take antiretroviral medication[75] and has been shown to impact on engagement with care[76]. Addressing stigma therefore is a key priority in HIV prevention, as recognised by UNAIDS, and tacking HIV stigma and discrimination in the UK is vital in ending the epidemic.

1.4 Clinical stages of infection

1.4.1 Natural history of HIV

HIV attacks host cells vital to immune function, such as CD4+ T-cells, macrophages and dendritic cells. The natural history and progression of HIV will vary widely between infected individuals, from rapid progression to long-term non-progression and long term viral suppression in 'elite controllers'. In general, shortly after infection, individuals may experience a brief early symptomatic period (seroconversion illness) followed by a period of clinical latency. Untreated, as the CD4 cell count depletes, usually after a period of years, patients will develop constitutional symptoms and eventually acquire AIDS defining illnesses, usually resulting in death within two years[77] (Figure 2).

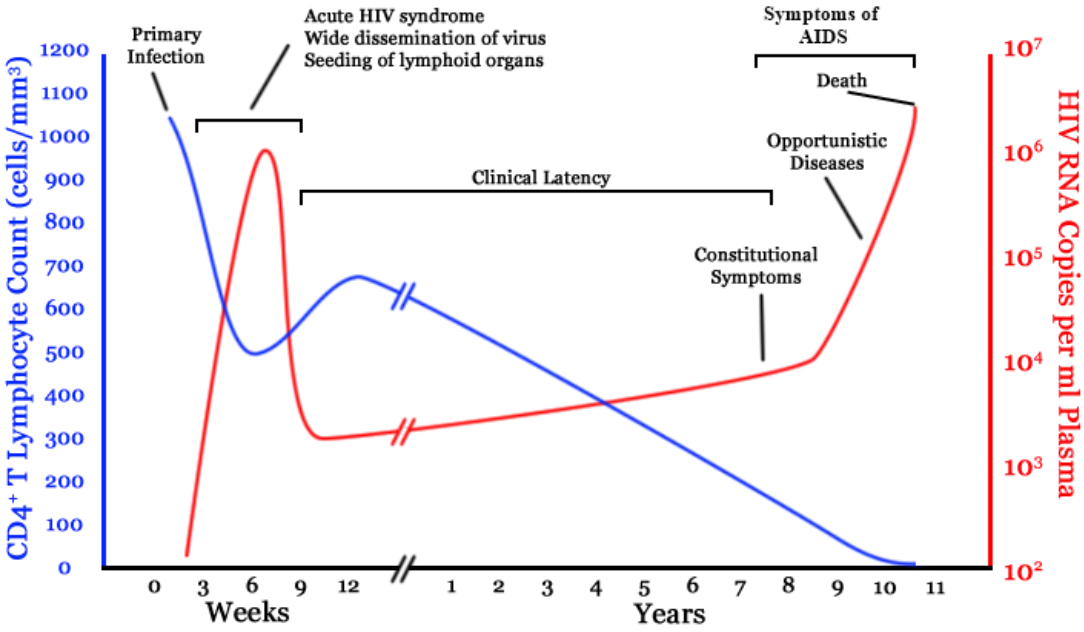


Figure 2: The stages and natural history of untreated HIV infection Reproduced from **Wikimedia Commons, GNU Free Documentation License**. By J Oliveira, and based on Pantaleo et al. 1993[78]

1.4.2 Staging systems

Staging of HIV infection is of importance both in the individual management of newly diagnosed infections, and in epidemiological monitoring and public health planning. The

two major classification systems used are the Centres of Disease Control and Prevention (CDC)[79] and the World Health Organization (WHO)[80] classification systems. Both require a positive HIV test result.

The CDC classification system bases staging on the absolute CD4 count and clinical symptoms, Stage 1 (CD4 >500 cells/mm^3) and stage 2 (CD4 200-499 cells/mm^3) being asymptomatic, and stage 3 categorised by either CD4 <200 cells/mm^3 or by the presence of an AIDS defining illness. In addition, stage 0 refers to early infection (diagnosis within 6 months of aquisition) irrespective of CD4 count. The WHO classification system uses primarily clinical staging (Primary HIV and stages 1-4), which may be supported by CD4 cell cut-offs where available, allowing for clinical decision making and management in resource poor settings.

1.4.3 Recent HIV infection

Recent HIV infection (often described as *early HIV infection*) is usually defined as infection diagnosed within the period of 6 months post acquisition. This includes the initial period post acquisition (primary or acute HIV) and the asymptomatic period following this.

Primary HIV, often termed seroconversion describes the period of development of HIV antibodies, which is symptomatic in 40-90% of infected individuals shortly after acquisition, and may be of varying duration and severity[81]. After successful transmission of HIV with local replication, the virus rapidly disseminates to regional lymph nodes, and virus may be detectable in the serum within 4-11 days[82]. Symptoms usually occur at 2-4 weeks post acquisition, with the highest prevalence corresponding with peak viraemia, approximately 2 weeks post detection of RNA in the blood[83]. Symptoms vary widely and are usually non-specific, commonly manifesting as fever, fatigue, myalgia, pharyngitis, rash, headache and diarrhoea[84].

Diagnosis of early HIV, including appropriate tests, will depend on time of potential acquisition. A high index of clinical suspicion is required in the very early stages, due to

the non-specific nature of symptomatic illness and may be guided by history of exposure. If recent negative test dates are available, a potential period of acquisition may be estimated. Viral RNA is the first detectable marker of HIV, being positive at around 10 days, followed by p24 antigen (included in standard 4th generation HIV tests) at approximately 14 days, enzyme-linked immunosorbent assay (ELISA) and western blot. In 2009, national RITA (Recent Infection Testing Algorithm) testing was introduced in England and Northern Ireland[85], incorporating avidity testing with ELISA, with CD4 cell counts, history of presentation with AIDS defining illness and ART exposure, to all newly diagnosed individuals[86]. Although the primary role of RITA is in surveillance of infection, a positive result is suggestive of recent HIV in the absence of clinical features of chronic disease[87].

Identification of recent HIV is of importance both for HIV prevention and individual clinical management. Recent HIV, due to an early rise in viral load[27] particularly in the acute stages, coupled with unmodified risk behaviour when undiagnosed, has shown to be a significant risk factor for HIV transmission[88, 89, 90], and as such, early diagnosis and treatment is a priority for prevention of transmission.Earlier diagnosis may also improve partner notification outcomes[64].

Early diagnosis of HIV additionally has significant benefits to individual health outcomes, with late presenters having a ten-fold risk of death within a year of diagnosis compared with those diagnosed early[91]. It also provides an opportunity for earlier intervention, with evidence showing initiation of ART at diagnosis irrespective of CD4 count has significant benefit in all-cause mortality and morbidity compared to initiation when the CD4 count drops below 350 cells/mm^3, as previous UK guidelines advised[51]. In addition, initiation in the acute stages of infection has been shown to reduce the viral reservoir[92], sites in which the virus accumulates and persists separately from the main pool of replicating virus[93], further reducing HIV associated complications and potentially dissemination to sanctuary sites, which may be of significance with the development of HIV cure strategies[94].

1.4.4 Asymptomatic and symptomatic HIV

Asymptomatic infection describes the period without symptoms in an untreated individual between recent HIV and the development of HIV related illness. Despite a lack of symptoms, HIV replication and depletion of CD4 cells continues, and constitutional symptoms may occur when the CD4 count drops below 500 cells/mm^3. The period of latency is directly related to viral load[95] and may vary widely between individuals, typically lasting six to eight years[96] though may be very brief in rapid progressors, or extended in long term non-progressors.

1.4.5 AIDS

AIDS is a clinical diagnosis characterised by the presence of one or more AIDS defining illnesses, including opportunistic infections and specific malignancies, for example, oesophageal candidiasis, pneumocystis jirovecii pneumonia and Kaposi sarcoma. These usually occur after the CD4 count has dropped below 200 cells/mm^3 but may present at higher counts. The term advanced HIV is often used when the CD4 count drops below 50 cells/mm^3 and untreated, has a median survival of approximately 12 months [97]. Late stage infection is associated with an increase in viral load, and therefore transmissibility.

1.5 Antiretroviral therapy and drug resistance

Since the discovery of Zidovudine's (AZT, a nucleoside analog reverse-transcriptase inhibitor, NRTI) activity against HIV in 1985, which was approved by the FDA in 1987, the idea that HIV was an untreatable killer was challenged. Unfortunately, with rapid development of resistance to this single agent, the clinical benefit of AZT alone was limited, with an increased death rate after two years when compared to untreated patients[98]. Dual therapy with newly introduced NRTIs showed some initial benefit[99], though the effects were again short lived. With the discovery of protease inhibitors and their inclusion in triple combination antiretroviral therapy[100] (also known as highly-active antiretroviral therapy – HAART), HIV has become a manageable, chronic condition, with

a near normal life expectancy, particularly in the context of earlier diagnosis [101, 102, 103].

The aim of ART is to supress viral replication and improve immune function. Treatment goals in the UK are to obtain an undetectable viral load (usually <50 copies/ml), reducing morbidity and the risk of developing drug resistance, alongside preventing transmission of HIV (section 1.4.3). Current UK guidelines recommend starting ART as soon as is practicable following diagnosis, with a regimen tailored to comorbidities, lifestyle, concomitant medications and baseline resistance if present. Most recommended regimens contain 3 active agents; typically a dual NRTI backbone, with a third agent (an integrase inhibitor (INI), non-nucleoside reverse transcriptase inhibitor (NNRTI) or a protease inhibitor (PI). Recently, with the introduction of integrase inhibitors, viral suppression may be achieved though dual therapy, for example with dolutegravir and rilpivirine or lamivudine (an INI, NNRTI and NRTI respectively), and monotherapy with boosted darunavir, a PI, may be used to achieve suppression in some cases.

Drug resistance to one, or several classes of drug usually occurs through inadequate viral suppression during antiretroviral treatment, particularly with NRTI and NNRTIs. This may be through historical use of mono- or dual-therapy with older agents or inadequate drug levels due to non-adherence, malabsorption or interactions with concomitant medications. Resistance may also occur in individuals who acquire HIV whilst using pre-exposure prophylaxis (PrEP) suboptimally. Due to rapid and error prone replication of the virus, viral replication in the presence of subtherapeutic drug levels may lead to the development of drug resistance mutations (DRMs) conferring resistance to the drugs it has been exposed to, and others within the same classes. Transmitted drug resistance (TDR) may occur when an individual with DRMs transmits the virus to a drug naïve individual. The prevalence of TDR in antiretroviral naïve individuals the UK was 7.3% for any type of mutation in 2014; 3.2% for NRTI mutations, 3.4% for NNRTI mutations and 1.8% for PI mutations[104]. Genotypic HIV drug resistance tests are therefore recommended for every individual at diagnosis and in the event of treatment failure in the UK[50]. These identify DRMs to antiretroviral drugs guiding possible treatment options. Standard baseline testing uses sequences from part of the *pol* gene to identify mutations reducing susceptibility to NRTIs, NNRTIs and PIs. Integrase resistance tests are available, but due to an apparent low level of TDR to integrase inhibitors, they are not currently recommended at baseline.

1.6 HIV virology

1.6.1 The discovery of a new virus

In 1983, two years after AIDS was first recognised in clusters of homosexual males, a retrovirus was isolated in Paris from the blood of an immunodeficient individual. Similar viruses were subsequently cultured by two parties in the United States, and named lymphadenopathy-associated virus (LAV), AIDS-associated Retrovirus (ARV) and human T-lymphotropic virus-III (HTLV-III) respectively[105, 106]. These were later found to be the same virus, which was subsequently renamed HIV in 1986[107].

1.6.2 Origins

HIV originates from zoonotic transmission from African primates infected with Simian Immunodeficiency virus (SIV)[108]. Two forms exist, HIV-1 and HIV-2 (see section 1.2.4) with distinct origins; HIV-1 originating in chimpanzees (*Pan troglodytes)*[109], and HIV-2 in sooty mangabeys (*Cercocebus atys)*[110]. Phylogeographic analyses have revealed the group M pandemic (the main group of HIV-1 viruses) originated in Kinshasa, the capital of the Democratic Republic of Congo (DRC) in the early 1920's, with an exponential growth rate in the early 1960s, possibly due to changes in commercial sex work and the use of non-sterilised injections at sexually transmitted disease clinics[111]. The dissemination of HIV globally appears to have occurred via a series of bottlenecks, explaining the predominance of various subtypes in different regions of the world. Subtype B, the commonest subtype in west and central Europe and the Americas, is the only subtype not found in Africa. It split from subtype D more recently than the separation of other subtypes and appears to have been introduced to the Caribbean in the mid-1960s, in connection with political changes in Haiti, initially appearing in a predominantly heterosexual epidemic. It then spread to homosexual men in New York in the 1970s before subsequent spread in this community through the USA and across to Europe[112]. Subtype C, the most prevalent subtype globally has been shown to have originated in Africa in the 1950's[113], spreading to India in the early 1970's[114] and from East Africa to

the UK in the early 1980's, with likely onward spread from the UK to Brazil[115].

1.6.3 Structure

(HIV) is a lentivirus of the family *Retroviridae.* Each spherical virion is roughly 120 nm in diameter, and comprises two copies of single stranded RNA, enzymes, including reverse transcriptase, integrase and protease, within a capsid enclosed by a viral envelope including the glycoproteins gp120 and gp41 which initiate binding and entrance into the host target cell.

The HIV genome is approximately 10,000 base pairs in length. This encodes nine genes including three major structural genes; *gag* encoding internal structure proteins, *pol* encoding enzymatic proteins, and *env,* encoding transmembrane proteins. In addition, two genes encode regulatory proteins (*tat, rev*) and four, accessory proteins (*nef, vpr, vif, vpu*). *Pol* encodes the reverse transcriptase (RT), protease (PR) and integrase enzymes, the targets of most of the common antiretroviral medications, and as such is sequenced for HIV drug resistance testing in routine HIV care.

1.6.4 Life cycle

HIV binds to the CD4 receptors of host T-lymphocytes through gp120, which then undergoes conformational change to allow additional binding to the host's co-receptors (CCR5 or CXCR4) causing membrane fusion. Once the viral contents are released into the target cell, reverse transcription occurs, where the viral enzyme reverse transcriptase converts the single stranded RNA to double stranded HIV DNA, which enters the host's nucleus during integration, catalysed by the HIV enzyme integrase, at which point the DNA is termed provirus. Transcription to mRNA is catalysed by the host enzyme RNA polymerase which is translated to viral proteins, cleaved into mature proteins by the HIV enzyme protease. This is followed by assembly of new virions within the host cytoplasm, and budding and release from the host's membrane of virions capable of infecting further cells (figure 3), though the virus may lay dormant within the host cells DNA, a phenomenon known as latency.

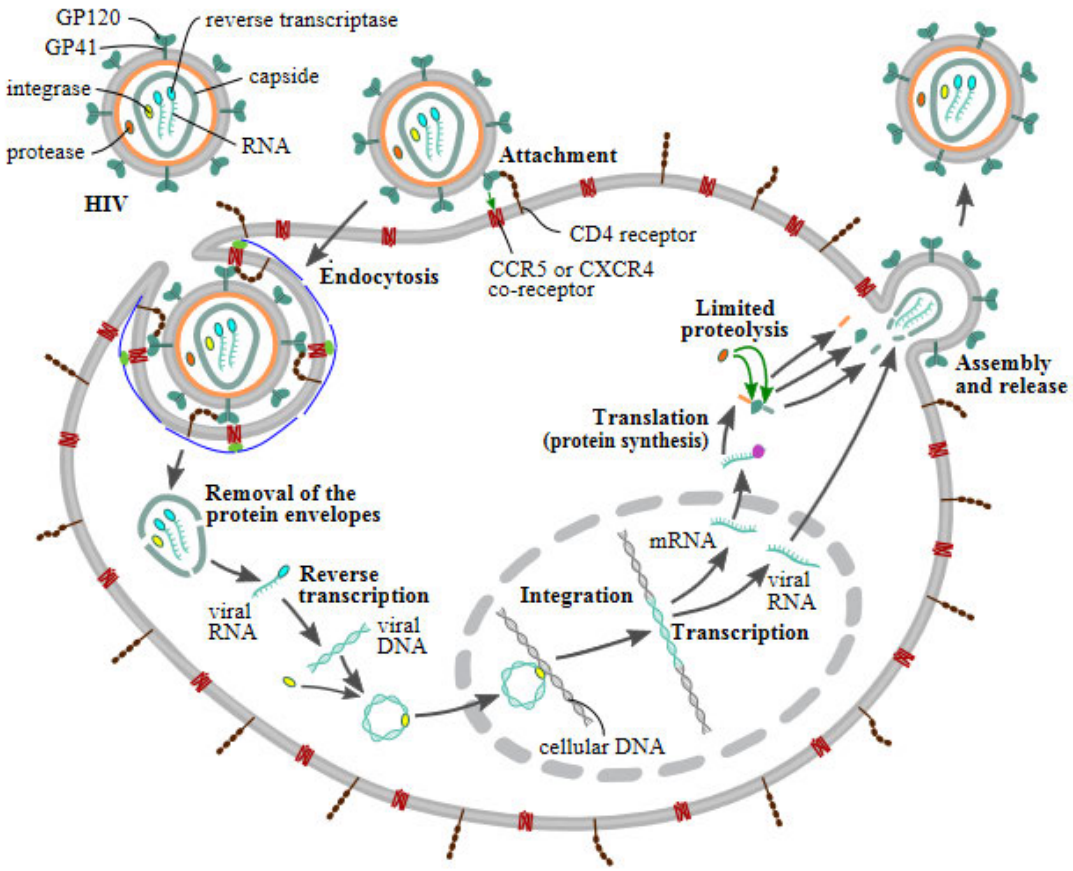


***Figure 3*: The HIV life cycle. By Jmarchn. Reproduced from Wikimedia commons. Licence CC BY-SA 3.0**

1.6.5 Viral evolution

During acquisition, the recipient is usually infected by a single, or a very small number of viruses, resulting in transmission bottlenecks[116]. HIV is a rapidly replicating virus, with a half-life of 1-2 days[117]. In addition, due to the error prone nature of reverse transcriptase during transcription of RNA to double stranded DNA an error is made every 1,000-10,000 bases[118], and two different copies are packaged within virions which may combine to form *recombinants*. This occurs during reverse transcription, where newly formed DNA can switch between the two original copies of RNA, and may occur anywhere along the genome, several times per replication and between different viral subtypes, leading to increased viral diversity[119].

Owing to this within-host evolution of the virus, diversity within an individual increases over time resulting in a population of different viruses, termed the quasispecies[120]. This high genetic variability allows the virus to evolve rapidly and evade the hosts immune

response, and in some cases develop resistance to commonly used drugs.

1.6.6 Classification

HIV-1 strains from different geographic regions have been found to phylogenetically cluster together into distinct groups, or clades, named A1, A2, B, C, D, F1, F2, G, H, J and K, letters referring to subtypes, with numbers referring to sub-subtypes (very closely related, but distinct lineages)[121] (figure 4). Collectively, these are termed group M ('main'), being the predominant subtypes globally. Group O ('outlier') encompasses phylogenetically distant strains, first identified in Cameroon[122], group N ('non-M, non-O' or 'new') unique non-M non-O subtypes[123] and group P, a rare and recently identified HIV strain closely related to gorilla simian immunodeficiency virus (SIVgor)[124]. Each of these groups represents a distinct zoonotic transmission from either chimpanzee (M, N, O) or gorilla (P). Circulating recombinant forms (CRFs), refer to recombinant linages with an identical mosaic structure.

HIV-2 is a distinct virus, which is closely related to sooty mangabey SIV, divided into distinct groups (A-H) due to the diversity of the virus, with very few CRFs.

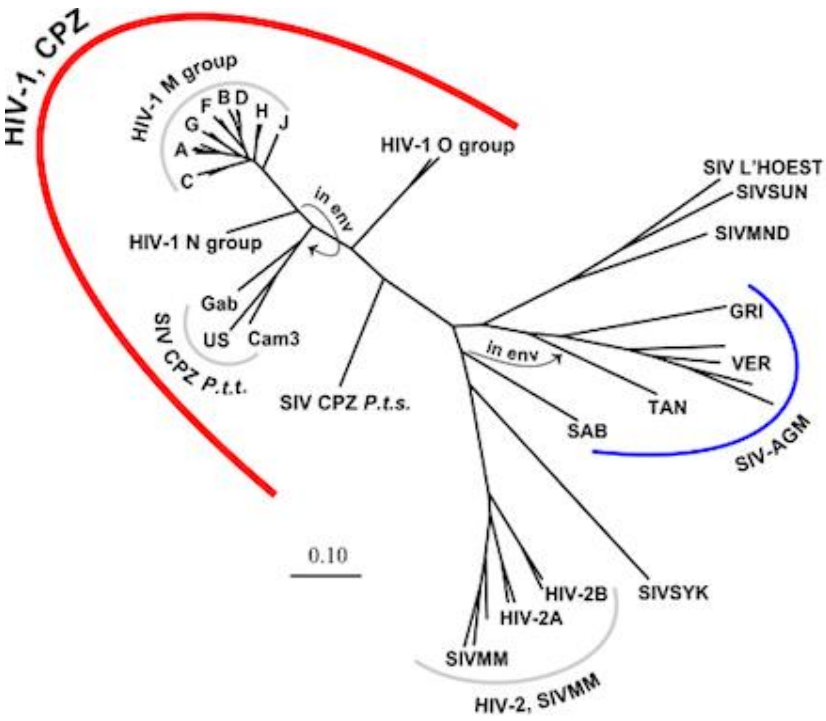


***Figure 4*: Phylogenetic relationships between HIV-1, HIV-2 and SIV based on the pol region. Scale = nucleotide substitutions per site. Reproduced from Kuiken et al. (1999)[125] via Wikimedia commons.**

.

1.7 Phylogenetics; methods, applications to public health and opportunities for prevention

Rapid evolution due to high replication rate, high mutation rate due to error-prone replication and recombination, combined with responses to selection from immunologic pressures and antiretroviral drug selection has led to enormous diversity of the HIV virus between both in and between individuals[126]. This diversity has been extensively studied using phylogenetics, the study of molecular evolution of an organism, a process of exploiting small differences in genetic or morphological features to determine the relatedness of organisms, assuming they are homologous (i.e. share a common ancestor)[127]. From these analyses, evolutionary relationships between genes or organisms

can be illustrated using a graphical representation, a phylogenetic tree or network map. In HIV, sequences that are highly related are joined in a cluster, which may be suggestive of recent transmission between them. Clustering patterns will depend on the population under investigation, the sampling density[128] and the phylogenetic methods and parameters used to recreate the network, for example a smaller genetic cut-off (a smaller genetic difference between sequences) will result in clusters of more recent transmission with a higher level of genetic relatedness.

1.7.1. Phylogenetic methods

Phylogenetic methods, and the use of phylogenetics has rapidly advanced since Darwin's first musings of phylogenetic relationships, drawn in a notebook in 1837 (figure 5), Understanding and technological advances have vastly expanded the power of phylogenetic inference and possibilities they provide. Despite advances, there are multiple stages and complexities that must be considered when using these techniques, and awareness of these processes is crucial to understanding their place within epidemiological analyses.

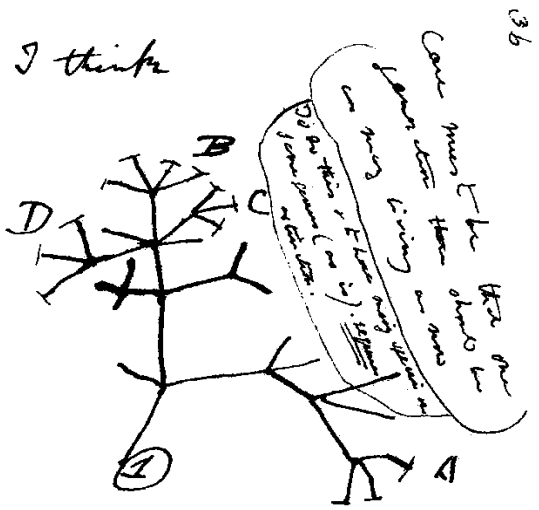


***Figure 5*: A sketch from Charles Darwin's notebook depicting an evolutionary tree, from around July 1837. Via Wikimedia commons.**

1.7.1.1 HIV sequences

RNA consists of a chain of nucleotides. Each nucleotide contains a ribose sugar with 5 carbon molecules. Attached to the first is one of four bases, two of which are purines; adenine (A) or guanine (G), and two pyrimidines; cytosine (C) or uracil (U) (which is replaced by thymine (T) in DNA). Each triplet of bases (a codon) encodes a specific amino acid which in sequence, are polymerised into a protein.

***Figure 6* : A section of a subtype B HIV-1 sequence**

All amino acids apart from tryptophan and methionine, which acts as a start codon for protein translation, have more than one triplet code. Frequently therefore, point mutations, particularly in the third codon, have little impact on protein structure, and are termed synonymous mutations. Transitions occur when a purine is replaced by a purine, and transversions when a purine is replaced with a pyrimidine or vice versa, the latter being more likely to impact on a protein structure than the former. Although there are twice as many possible transversion errors (A↔C, A↔T, G↔T, G↔C) than transition errors (C↔T, G↔A), the more disruptive transversions tend to occur less frequently due to biochemical and steric effects[127].

The most frequently used section of the HIV genome for phylogenetics is part of the polymerase (*pol)* gene, processed for the purpose of HIV drug resistance testing derived from Sanger sequencing, as this encodes for the targets of the commonly used classes of antiretroviral drugs. Within an individual, multiple different viruses exist due to intrahost evolution, sometimes termed the quasispecies after Eigen and Schuster[129]. Sanger sequencing produces a consensus sequence of the most common nucleotides within the viral quasispecies, which will identify resistance occurring of frequencies above approximately 20%. Although the pol region is relatively conserved, it has been shown to provide sufficient information for investigation of HIV transmission[130].

Next-generation deep sequencing (NGS) uses high throughput methods to sequence a larger proportion, or all of the available clones within a sample, with multiple short reads per virus, which can provide a more accurate representation of the quasispecies. This can

detect resistance mutations present at a much lower level, though the clinical significance of low level resistance mutations is unclear[131]. Although high throughput methods for NGS mean sequences can be created both more quickly and at a lower cost, it is currently not used for standard resistance testing though is being utilised more widely within the research setting. Viruses may be directly sequenced for the purposes of phylogenetics, or retrieved from a central database. Examples include the UK HIV Drug Resistance Database, (UK RDB) a central repository for routine resistance tests in the UK, and open access repositories, such as Genbank or the Los Alamos National Laboratory HIV Sequence database (LANL), where sequences can be searched according to subtype, genetic region of the sequence or demographic factors such as country of infection.

1.7.1.2 Sequence alignment

Prior to analysis, nucleotides must be aligned in-frame (as codons) so equivalent positions are compared (positional homology). This is often performed against a reference sequence, commonly for HIV-1, the HXB2 sequence. Each nucleotide position is classified as a match, mismatch or a gap, which represents an insertion or deletion (collectively referred to as indels) or combination of the two, within either of the sequences being aligned. This may be performed in a pairwise fashion between two sequences, or through multiple sequence alignment, usually through a progressive alignment algorithm. Computerised software is available such as ClustalX[132] or T-Coffee[133] which rapidly align chosen sequences though this should be followed by a manual check and edit, using software such as BioEdit[134] or Geneious[135], prior to phylogenetic reconstruction.

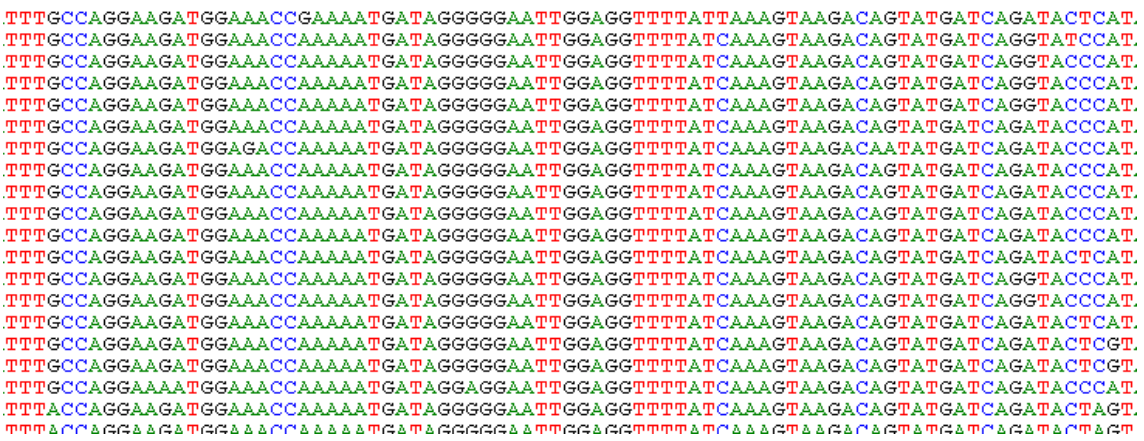


***Figure 7*: Aligned subtype B HIV-1 sequences**

1.7.1.3 Nucleotide substitution models

Once a set of sequences is aligned, a nucleotide substitution model is specified which converts genetic distances (the proportion of differences between nucleotides within sequences) into presumed evolutionary distances, using assumptions for nucleotide base frequency, rates of substitution, frequency of transitions vs. transversions and variation between substitution rates across the DNA. The simplest model is the Jukes Cantor (JC69) model[136], which assumes all nucleotide frequencies are equal, and during evolution, each is equally as likely to be replaced by any other. The Kimura 2-parameter (K80) model[137] also assumes base frequencies are equal, though allows for the fact that transitions occur more frequently than transversions. The Felenstein (F81) model[138] allows an unequal frequency of base substitutions, and the Hasagawa, Kishino and Yano (HKY85)[139] model allows both the nucleotide frequency and rate of transitions and transversions to differ. The General Time reversible model (GTR)[140] is the most complex model, and incorporates different rates for every change and allows differing frequencies of nucleotides. In addition, a proportion of invariable sites (+I)[141] and/or rate of variation across sites (+G)[142] may be added into any model. The GTR +I+G model usually fits real life data most accurately[143], and is the most commonly used model in HIV, where more complex models are generally preferable[144, 145].

The most appropriate evolutionary model can be identified using software such as ModelTest[146] which uses the sequence data itself to select the model with the highest likelihood score. A graphical representation, a phylogenetic tree or network map is then constructed using the aligned sequences.

1.7.1.4 Molecular clocks

Rates of divergence allow us to allocate timings of nodes within trees. This is referred to as a molecular clock. The molecular clock hypothesis was developed in the early 60's after it was observed that the genetic distance between sequences encoding the same protein increased linearly with time[147]. This 'strict molecular clock' hypothesis states that all lineages evolve at the same rate, allowing for estimation of the date of internal nodes and roots of the tree. Clocks can be 'calibrated' using historical samples, and dates of

divergence may then be estimated from sampling dates. Later, Felsenstein proposed the 'relaxed' unconstrained model, where each branch has its own rate of evolution independent of all others[138]. More recently it has been accepted that these two extremes are biologically unrealistic, and an 'uncorrelated relaxed molecular clock' hypothesis has been proposed[148], which incorporates the time dependent nature of evolution without the need for a strict clock, and can estimate divergence times based upon the mean clock rate of the whole tree, and vary according to the underlying distribution of the data.

1.7.1.5 Phylogenetic trees

Phylogenetic trees provide a graphical representation of the relationships between taxa (the population being investigated), providing insight, for example of the evolution and spread over time (section 1.6.2). Each sequence is represented by a 'tip' of the tree, and are linked by branches to internal 'nodes', representing a common ancestor. More than two branches linked by a node are defined as a clade, which may be classified as a cluster. Trees may be unrooted, or rooted (all linked to one initial node, or common ancestor) which as well as providing information on genetic linkages, provides a scale, with branch lengths representing evolutionary distances between sequences. The branching pattern is referred to as the topology.

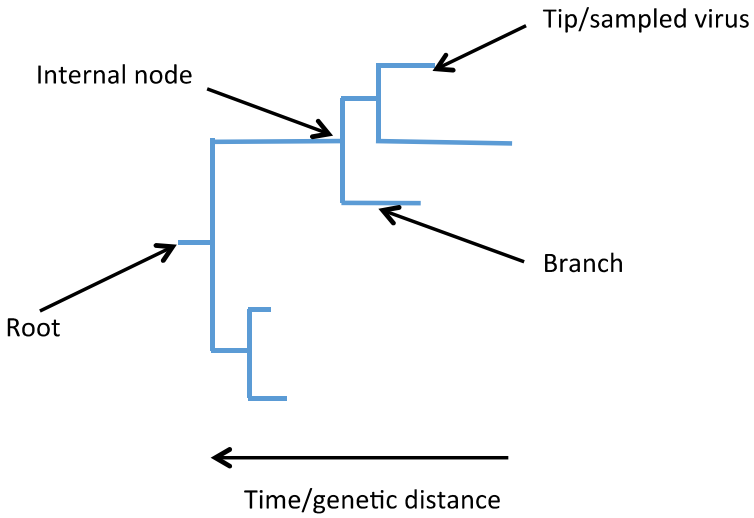


***Figure 8*: A phylogenetic tree**

There are several methods for constructing trees, the most appropriate being determined by the sequence data. The methods fall into one of two categories; distance-based methods and character-based methods. Distance based methods calculate the genetic distances (values based on the fraction of positions in which two sequences differ) between every sequence under analysis, creating a distance matrix which is then used to construct a tree. The unweighted-pair group method with arithmetic means (UPGMA) or the weighted-pair group method with arithmetic means (WPGMA) are two distance based methods that assume the rate of evolution is constant, a theory which is no longer accepted which has led to a decline in their use in phylogenetics. These methods use a clustering algorithm to build a tree in a stepwise manner, by sequentially combining sequences with the smallest genetic distance. The neighbour joining (NJ) method[149] works in a similar way, but creates the tree according to the 'balanced minimum evolution' criterion; rather than clustering the most closely related sequences, it attempts to minimise the length of internal branches, and therefore also the length of the overall tree. Unlike UPGMA and WPGMA, NJ does not assume the same evolutionary rate for all lineages, and allows for an evolutionary model. Being computationally quick and able to analyse large volumes of data, neighbour joining is now the most commonly used distance method, however it has been known to produce inaccurate trees, and like other distance methods, information on individual sites from sequences is lost.

Character based methods, including maximum likelihood, Bayesian, and maximum parsimony use information directly from sequences, rather than relying on genetic distances alone. These methods are generally more computationally demanding but provide improved accuracy in representation of phylogenetic clustering by applying a statistical model of evolution. Maximum likelihood methods calculate the probability of a tree according to the patterns of sequence alignment, given a specified substitution model, based on likelihood probabilities[150]. Likelihood (L) is proportional to the probability (P) of the observed data (D), given the hypothesis (H);

$$L=P(D|H)$$

One problem with this method is that the number of trees produced to find that with the maximum likelihood increases massively with the number of sequences analysed, which can limit the number of trees that can be examined. Heuristic strategies can be applied in

these cases, often using a neighbour joining tree as an initial topology, with likelihood estimates applied to subsequent rearrangements. Bayesian methods require the specification of a prior belief, or hypothesis (prior distribution) which is often vague where little is known about the data, from which a posterior probability is calculated, according to Bayes rule, which calculates the probability of the hypothesis (prior) after the data is observed;

$$Pr(H|D) = \frac{Pr(D|H)Pr(H)}{Pr(D)}$$

Where Pr(H|D), the posterior probability of the specific hypothesis (H), given the observed data (D) is equal to Pr(D|H) (the conditional probability of observing the data if the hypothesis is true) multiplied by Pr(H) (the prior probability of the hypothesis, inferred before observing the data) divided by Pr(D) (the marginal probability of the data – the probability of observing the data under all possible hypotheses.)

Practically, trees with the highest likelihood are selected for the given prior distribution and a posterior probability is formed by exploring tree space and using the Markov Chain Monte Carlo sampling method, which randomly searches for a tree with the highest likelihood and compares it to another randomly generated tree. If the likelihood of the next tree is higher, this is accepted, if not, the first tree is held in memory. The more times a tree is held in memory, the higher its likelihood. Unlike other tree construction methods, Bayesian phylogenies provide an inbuilt measure of statistical support, removing the need for additional processes such as bootstrapping (see below). Each tree topology that is well supported by the data is combined to produce a maximum clade credibility (MCC), or consensus tree for which each branch is allocated a probability.

1.7.1.6 Statistical support

Bootstrapping[151] is the most common method used to infer the reliability of the clades within a tree. It is a non-parametric resampling method which calculates statistical error where the underlying sampling distribution is unknown. This method can be applied to all tree construction methods to provide a measure of tree robustness, though the use of posterior probabilities in Bayesian analyses means bootstrapping is not required. In this method, a new alignment is obtained from the original sequence data by randomly

choosing columns with replacements. Each column can be selected multiple times or not at all until a bootstrap replicate is obtained (a new set of sequences, the same length as the original). For each replicate, a new tree is constructed, and the proportion of matching trees is calculated, giving a bootstrap value as a measure of the monophyly of the subset. Bootstrap values of ≥70% are generally considered to indicate strong support for a clade, however, high bootstrap values suggest there is no other very close relative to the clade, not necessarily that the members of the clade are very closely related themselves[152], and have been omitted in investigations aiming to identify close transmission partners for this reason[153].

1.7.1.7 Phylogenetic networks

Phylogenetic trees are particularly useful for differentiating between distinct lineages, however within clades, identifying transmission partners, or very closely related sequences is difficult. In addition, one internal node within a tree can only be linked to two branches, which may not clearly represent the patterns of transmission in HIV, where several individuals may have acquired the virus from one other. Transmission networks aim to directly link closely related sequences, more clearly demonstrating similarities between pairs, or groups of sequences. Identifying 'transmission pairs' may be difficult within clusters from phylogenetic trees, whereas in a network approach linking very closely related sequences, the internal structure of a cluster is maintained. Network approaches frequently use a distance approach to identify these closely related groups, which is much less computationally demanding, allowing for larger scale, even global analyses[153].

1.7.1.8 Defining a cluster

Previous studies have identified clusters as subtrees (or clades) linked by an internal node, the most recent common ancestor (MRCA) that have appropriate statistical support after an initial tree is inferred. Clustering may be defined according to node support, genetic distance or a combination of the two. This will be guided by the questions asked of the data, and by the dataset itself[145]. Frequently, when investigating recent or rapid transmission, shorter genetic distance cut-offs are used, however this may reduce large

and long-lived clusters down to multiple sub-clusters, losing valuable epidemiological information. Software can be used to identify clusters, including ClusterPicker[154], PhyloPart[155] and HIV-TRACE[156]. These all identify clusters under a given genetic distance threshold, though HIV-TRACE identifies clusters much faster when using larger datasets.

1.7.2 Limitations of phylogenetics

Though useful insight may be gained on a population level with phylogenetics, limitations mean we cannot confirm direction or directness of transmission on an individual level with phylogenetics alone, for several reasons. These limitations are due to both practical and phylogenetic issues.

Firstly, realistically, it is highly improbable that any phylogenetic analysis will include sequences for every individual within a network. A proportion of transmissions will occur from individuals who are, and may remain, undiagnosed for some time, from whom sequences will obviously be unavailable. Even within the diagnosed population, a proportion will not have sequences available for several reasons; firstly they may have started treatment prior to sequencing, particularly if diagnosed before routine resistance testing became part of practice. Secondly, some viruses fail to sequence, particularly if the viral load of the sample used is low. Some patients will be diagnosed abroad, where sequencing is not part of routine practice, or if done, these may be difficult to retrieve, and others may have been recruited at diagnosis to clinical trials, with sequencing performed within a study, which will unlikely be available to outside researchers for phylogenetics. Within the UK, the UK DRB hold sequences for only 89% of the current diagnosed population.

Simulation studies have shown the detectability of HIV clusters to be substantially affected by sampling density, with densities of <10% associated with stochastic clustering and broad confidence intervals[128]. A genotyping density of >10%, with a sampling density of 50-70% has been recommended for HIV cluster analysis, though evidence from a global analysis of all available pol sequences (84,527) revealed not only expansion of clusters identified in previous localised phylogenetic studies, but identified several

clusters containing sequences from different countries, suggesting that sampling, even at a high density from a single region or country will likely underestimate clustering[153].

Due to incomplete sampling therefore, within any one analysis it is usually impossible to conclude with certainty that within a transmission pairing there cannot have been any unsampled intermediaries, thus direct transmission cannot be inferred from phylogenetics alone.

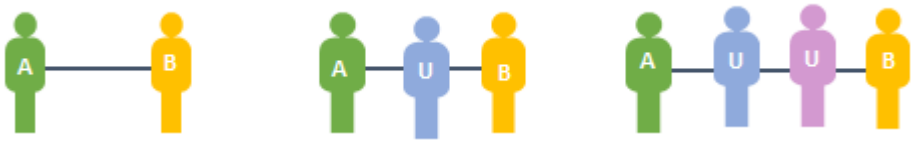


***Figure 9*: Possible transmission events involving 2 closely linked sequences, A and B. U= unsampled**

In addition, phylogenetic analysis using Sanger sequences from routine resistance tests cannot infer direction of transmission. Supportive clinical information however may suggest direction, including dates of positive and negative testing, blood tests suggesting seroconversion, such as evolving serology or western blot, and to a degree, other clinical information suggesting duration of infection such as AIDS diagnoses, CD4 T cell counts and RITA testing.

1.7.2.1 Advances in determining direction/directness

In recent years researchers have developed methods that may infer direction or directness of transmission, with the use of multiple sequences from individuals, or the use of next generation sequencing with multiple clones from each individual's sample. Transmission of multiple lineages of HIV occurs in around 20-40% of transmissions[157, 158] and by creating phylogenetic trees containing multiple clones for each individual, inferences can be made from the patterns of branching, and mixing between different individual's sequences. This was explored in a simulation by Romero Seversen[159], who postulated that direction can be inferred if certain branching patterns were observed. It was hypothesised that direction of transmission can be predicted if a paraphyly exists (one individuals sequences stem from another's, mixed into the tree – see figure 10), intermediaries may

be excluded if multiple lineages are transmitted from one individual to another and sequences are heavily intermingled, and when two individuals both have a monophyletic population (a group of descendants from one common ancestor), a common source is likely. Results from known transmission chains have supported these conclusions.

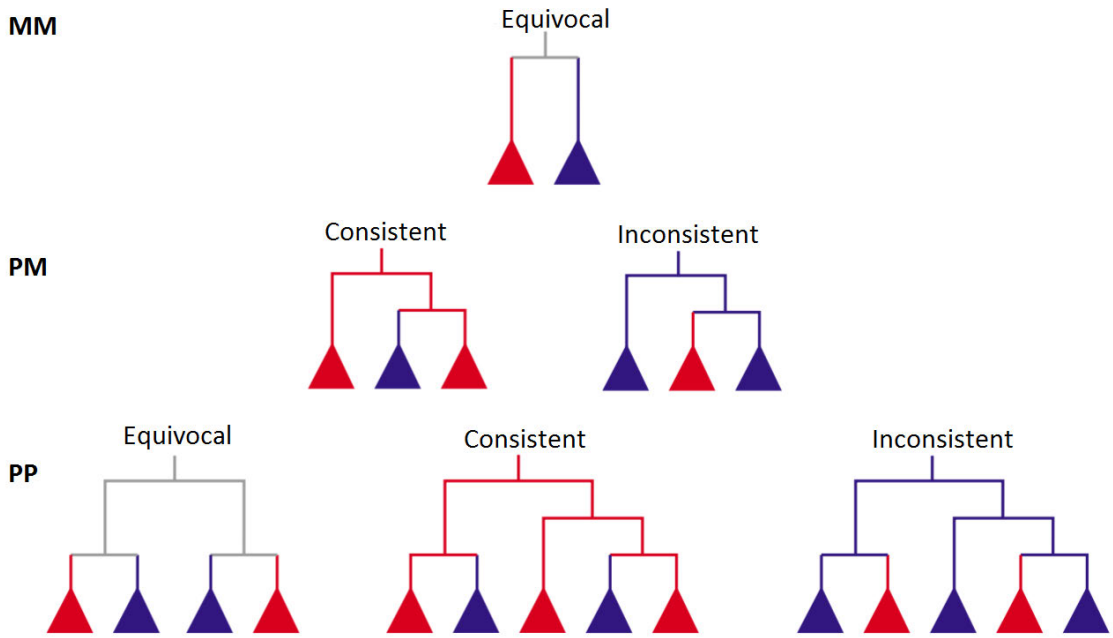


Figure 10: **possible topologies of transmission pairs. Exploring whether 'red' transmitted to 'blue'. MM= monophyletic-monophyletic. PM= paraphyletic-monophyletic. PP= paraphyletic-polyphyletic. Reproduced from Romero-Seversen et al. (2016)[159]**

Figure 1.8 shows potential topologies when exploring whether 'A' transmitted their virus to 'B'. MM topologies, where both populations are monophyletic do not provide information on direction of transmission, but are typical of a common source to both A and B. In PM topologies, one individuals set of sequences clusters together, but these sit within, and are stemmed from the other sequences. PM topologies may be derived from direct or indirect transmission but suggest transmission from the paraphyletic population to the monophyletic population. A 'PP' topology, where one individuals sequences are intermingled with the other's, suggests direct transmission. In a PP 'equivocal' topology direction cannot be determined, however a 'PP consistent' scenario, where the source's HIV population (A) is paraphyletic and recipient B is polyphyletic, supports direct transmission from A to B with near certainty.

Recent data using viral NGS sequences has supported this hypothesis to a degree. Data from confirmed transmission pairs within the HPTN 052 study was used to determine how confidently direction could be inferred using similar methods[160]. Given these data were collected within a clinical trial with close monitoring of blood tests, frequent HIV testing and documented behavioural data, this is a near perfect dataset to test this hypothesis. Results showed that direction could be inferred in up to 74%, meaning although these methods may provide useful insight on a population level, they cannot be used to confirm direct transmission between any two individuals. Similarly, a large study using NGS data from PANGEA recently explored whether direction, or directness could be determined using population viral sequence data. Similarly, there was an error rate of 16.3% (8.8-28.3%) when assessing direction of transmission[161]. Directness of transmission could be inferred in some cases, however two pairs with a PP topology, suggesting direct transmission from one to another were found to be both female gender, which would be a highly improbable within the study population, therefore similarly showing directness cannot be 100% confirmed at an individual level even using next-generation techniques.

1.7.3 Applications of phylogenetics in HIV

The relative small size of the HIV genome, along with short generation times and high substitution rates provide an ideal model for studying its evolution through phylogenetics.

The application of phylogenetic analysis in HIV began in the 1990s, initially being used to identify the origins of HIV, and to classify it into its types (1 and 2) and subgroups[109]. As well as identifying the origins of the HIV epidemic, phylogenetic analysis of HIV has been used to successfully analyse the molecular evolution of HIV, including the relationship between disease progression and rates of substitution[162], mutation selection dynamics both within and amongst hosts[163, 164], the discovery and understanding of recombination in HIV [165, 166], and characterisation of viral reservoirs[167, 168].

1.7.3.1 Uses in public health

1.7.3.1.1 Risk groups and sexual networks

Valuable information on contact networks can be derived from traditional epidemiological methods such as contact tracing in HIV (Section 1.3.6), but this may be poorly described using interview data due to factors such as recall and social desirability bias, and, particularly with chronic infection, does not necessarily equate with any accuracy to the actual transmission network. Despite its limitations, phylogenetic cluster analysis combined with clinical, epidemiological and geographical data can provide valuable insight into HIV transmission networks which contact tracing cannot provide. Phylogenetic studies of localised epidemics have shown mixing between different ethnicities and risk-groups, and also between heterosexuals and MSM unreported during clinical consultation[169, 170, 171], with suggestion that significant mixing between heterosexuals and MSM has driven transmission in certain cohorts, and heterosexual transmission alone is not self-sustaining[172]. Analysing risk groups may also help identify under-diagnosed groups. For example, large studies in the UK and USA have revealed clusters comprising of only women[171], which may reflect problems with recruitment or reluctance to test, perhaps due to a self-perception of lower risk in heterosexual men. On a larger scale, analysis of the *pol* gene from all available sequences worldwide has demonstrated large international clusters containing isolates from several countries, highlighting the role of international transmission, frequently underestimated as many phylogenetic analyses focus on one region or country[153]. Additionally, size of clusters have provided insight into the dynamics of sexual networks; larger clusters have been shown to expand disproportionately compared to smaller clusters, individuals within these clusters reporting larger social/sexual networks, those in smaller clusters reporting more high risk behaviour with regular partners[173].

1.7.3.1.2 Outbreaks

Phylogenetic analyses have also been used to investigate HIV outbreaks. One early example was that of the ‘Florida dentist case’ in the early 90’s, where 6 patients at the dental practice of Dr David Acer became infected. A phylogenetic analysis performed at the time indicated 5 of them were linked but the route of transmission was never

established[174], and subsequent presenting evidence has questioned the hypothesis that the mechanism for these transmissions was through dental procedures[175]. Phylogenetic analyses have supported the hypothesis of a single heterosexual source of a HIV-1 outbreak in Doncaster, UK[176], facilitating a co-ordinated public health response[177] and identified a higher than expected rate of within-prison transmissions amongst drug users in Glenochil, Scotland[178], prompting drug harm reduction measures throughout prisons in the country. On a larger scale, phylogenetic analysis identified migration and poverty underlying two large outbreaks in Greece and Romania, owing to increased needle sharing, guiding the resultant public health response[16, 179]. Recently, phylogenetics have been crucial in unravelling two large and unexpected outbreaks amongst people who inject drugs in Glasgow, UK, and Indiana, USA[17, 180].

1.7.3.1.3 Factors driving transmission

As phylogenetic clusters reflect groups within which multiple transmissions have occurred, integrating clinical and epidemiological data provides not only an insight into networks at risk of transmitting, but identifies specific clinical factors associated with transmission, providing a guidance for public health strategies. Primary/early HIV is one such factor that has frequently been implicated and there is substantial evidence to support the hypothesis that this can be a significant driver in the HIV epidemic[88, 90, 181, 182], one analysis in London showing that, within clusters of 10 or more individuals, 25% of transmissions occur during the first 6 months of infection[183]. As well as recent infection, higher viral loads[88, 171, 184] presence of an additional sexually transmitted infection [88, 185] and higher CD4[88, 171] (potentially related to more recent infection, or the deferral of ART due to previous clinical guidelines) have all been found to be associated with transmission though phylogenetic studies, reinforcing the importance of strategies including early ART for treatment as prevention. Phylogenetic analyses have also been crucial in providing evidence to support U=U, being used to support or refute transmission between partners in studies such as PARTNER[32], an observational study investigating transmission between serodiscordant partners.

1.7.3.2 The use of phylogenetics to guide real time interventions

Until very recently, phylogenetic analyses have been performed retrospectively, owing to processing of sequences and data, and time taken to conduct phylogenetics. Most inferences gained therefore have been used to guide future health responses. A phylodynamic analysis performed in the UK has shown that networks of MSM are likely to be characterised by preferential attachment (where individuals with more connections are likely to gain even more connections at a higher rate), and therefore randomly distributed prevention strategies are unlikely to disrupt on-going transmission[186]. As early as 2009, the use of phylogenetic analysis to directly guide public health responses by targeting individuals within highly connected networks was proposed, following a study designed to assess the utility of real-time voluntary partner notification in identifying phylogenetic clusters in recently infected patients using a transmission network of *pol* sequences[187]. Although there was a low success rate (<10%) of identifying and enrolling epidemiologically linked patients, it was found that over 60% of linked partners subsequently clustered within highly related transmission groups, concluding identifying transmission clusters may be useful for future prevention strategies focussed on these risk networks.

Using this same patient cohort in San Diego, USA, an observational study was subsequently conducted to determine whether baseline transmission networks could prioritise individuals for antiretroviral therapy to reduce onward transmission[188]. Within this study, newly diagnosed individuals were recruited, and data collected on phylogenetic network connectivity, baseline CD4, VL, behavioural data, stage of infection and demographics. A transmission network score (TNS) was allocated according to baseline phylogenetic network connectivity (a higher number of links, or edges correlating with a higher TNS). Patients were followed up for at least one year and TNS and baseline characteristics were correlated with number of onward transmission events, as suggested by additional phylogenetic connections after 1 year, to determine its validity as a prediction tool. Network connections at baseline (as quantified by the TNS) were found to be correlated with risk of on-going transmission and baseline risk behaviours (number of sexual partners and unprotected anal intercourse), with a reduction in transmission risk in those with a high TNS on effective ART, suggesting phylogenetic analysis may be a useful tool to target individuals for early treatment as prevention.

Only recently however, has the prospective use of phylogenetics been implemented to directly target individuals for public health intervention. In British Colombia, Canada, the use of phylogenetic analysis of baseline HIV drug resistance was used prospectively to guide public health response through the use of an automated phylogenetic system, where new sequences were added to the database as they were processed leading to an automatic reanalysis. Time from diagnosis to receipt of HIV sequences to the database was reduced to as little as 6 days meaning surveillance occurred in near real-time. In 2015, individuals within a rapidly expanding phylogenetic cluster containing two NNRTI mutations, suggesting multiple recent transmissions, were traced back and targeted for interventions including re-engagement into care, ART adherence support and repeat contact tracing[189]. Although there was no control group, subsequent analysis showed that transmissions that occurred after the intervention were mainly non-resistant viruses, and as only individuals with resistance were targeted for intervention, this suggests a public health impact from these actions.

Subsequent to this, the Centres for Disease Control (CDC) have implemented prospective phylogenetic analysis to guide public health interventions throughout the USA[190]. Rapid identification and response to growing clusters of infection has been identified as one of the 4 key pillars of the recent strategic initiative in the US to end the HIV epidemic over the next 10 years[191] making this area a key priority for prevention. This has been implemented using HIV-TRACE[156], a user friendly interface which uses genetic distance clustering to rapidly identify clusters using large datasets. Given resource limitations, efforts are focussed on identifying clusters of *recent and rapid* transmission, which are identified using tight genetic distance cut-offs. Their research has shown that 'priority' clusters with a genetic distance cut-off of 0.5%, identified as having ≥5 new diagnoses in the past 12 months have a transmission rate of 33 events per hundred person-years, compared with a national average of 4 events per hundred person years[192]. On identifying clusters, interventions that may be applied include engagement into care and treatment, partner notification and allocation of PrEP to contacts, alongside social network approaches. Although little data on outcomes has yet been published, the CDCs cluster detection has identified over 145 'priority' clusters to date, which has led to public health campaigns, triggered local public health responses, and averted the need for outbreak

investigation by demonstrating non-clustering amongst an apparent outbreak identified though a rise in incidence in a low prevalence area[190].

1.8 Should real-time phylogenetically guided strategies be implemented in the UK?

Given the wealth of available data in the UK, with a high coverage of resistance testing, real-time phylogenetic techniques may provide useful opportunities for unravelling 'outbreaks' and tackling the epidemic. Although recent significant achievements have been made in reducing incidence in the UK, the recent outbreak in Glasgow provides an example of where phylogenetic surveillance may provide benefit. Although this outbreak was initially identified though traditional methods – the observation of a series of routine genotypes with the same unusual subtype and resistance pattern – a subsequent phylogenetic analysis has revealed the extent of the cluster, containing over 100 individuals, which had been expanding for several years[17]. Phylogenetic surveillance would have likely identified this cluster earlier, allowing for public health measures to be appropriately applied reducing the extent of spread, which may have remained unnoticed had they been a more common subtype, without the specific drug mutations that were first noticed by staff in the local reference laboratory. Although this outbreak is exceptional within the UK epidemic, other smaller outbreaks amongst PWID have been detected in the South West, Dublin and Wales in recent years[193, 194], and with significant cuts to public health funding further outbreaks could be possible. In addition, as our treatment and prevention strategies improve, more targeted interventions to detect pockets of transmission may become more necessary.

The use of real-time phylogenetics to guide interventions however is an area in its early stages of development; research is on-going into the development of optimal metrics to monitor growth of clusters after detection and during intervention, and the prevention benefit of these measures is yet to be proven, so caution is required when considering their implementation in the UK. Given the time taken to process sequences ready for phylogenetic analysis, and recent changes in the UK epidemic, whether discreet targets

amenable to intervention exist is unknown. Given the uncertainties in terms of risks and benefits, an ideal introduction would be to locally pilot phylogenetically driven interventions to ascertain public health benefits, logistical and cost issues, and patient centred barriers and facilitators to ensure benefits outweigh risks, before its use becomes more widespread.

In this programme of doctoral research, I will therefore address the following questions;

1. How could phylogenetics be implemented in a way that is acceptable to patients in the UK?
2. Can real-time phylogenetically guided interventions be piloted on a local scale, providing evidence for effectiveness and real-life acceptability?

Chapter 2. A systematic review and narrative synthesis of the use of molecular epidemiology for clinical case finding in HIV and other stigmatised infections.

Chapter purpose and summary

Owing to the recency of developments in using phylogenetics to guide direct, near real-time public health interventions in HIV, potential strategies, acceptability and ethical issues may not yet have been fully considered. This chapter systematically explores previous experiences of case finding guided by phylogenetics, an intervention chosen given its additional complexities, with the involvement of additional contacts, compared to other potential interventions such as targeting individuals for early treatment. Given its more limited use in HIV, we sought to identify experiences where phylogenetics has guided case finding not only HIV, but other similarly stigmatised infections. The aim of this chapter is to identify methods, barriers, facilitators and ethical issues arising with previous use of phylogenetics for case finding interventions, in order to direct the subsequent body of research.

2.1 Abstract

Background
Phylogenetic information provides new horizons for clinical case finding in HIV, but raises issues of acceptability, privacy and even criminalisation. Studies describing use of molecular epidemiology to directly inform case finding in stigmatised non-nosocomial infectious diseases were reviewed to identify methods used, barriers, facilitators and ethical issues arising.

Methods
A search in MEDLINE, Embase, CINAHL and PsychINFO for articles where phylogenetics have been used to directly facilitate case finding in sexually transmitted infections, TB, HBV or HCV, published until May 2018 in English.

Results
38 of 7,063 papers screened met the inclusion criteria; 31 TB, 7 HIV. Case finding strategies included identifying or confirming outbreak cases instigating further investigation, epidemiological findings linked to molecular epidemiology used to guide the development of a targeted screening programmes and using linked epidemiological data from phylogenetic analysis to provide clues to a potential international transmission source.

Barriers included delayed results and investigations, cost and human resource required, reluctance of individual to name contacts and refusal of access to premises for screening. Facilitators included sharing molecular surveillance data to establish community support in targeted TB screening. Ethical issues included consideration of the release of an HIV sources identity to the media and weighing of individual risks against public health benefits.

Conclusion
Phylogenetics-informed approaches to case finding have been used in stigmatised infections to detect previously undiagnosed infection. However, studies reporting their use in clinical and public health practice provide limited information on patient related barriers, acceptability, or on ethical challenges such as identification of potential sources of infection or criminalisation.

Research into patient views on acceptability, risks and preferred approaches to using phylogenetic information for case finding in HIV is needed to inform future interventions.

2.2 Introduction

Undiagnosed HIV infection remains a significant issue in HIV prevention and control. As well as contributing to potential late diagnoses, undiagnosed infection has been estimated to account for up to 82% of new infections in men who have sex with men (MSM)[195], with the majority of the infectious population being in the undiagnosed category[33]. Current methods of identifying undiagnosed HIV infections include expanded testing, pre-exposure prophylaxis and contact tracing of recently diagnosed infections. Although a combination approach has reduced incidence in MSM in some areas in the UK, there may be limitations when trying to target harder to reach groups, a recent extensive outbreak of HIV in people who inject drugs in Glasgow exemplifying how unexpected transmission may occur despite intensive prevention strategies[196].

Phylogenetic analysis has been used for many purposes in public health investigations of HIV, including identification of risk groups[197], risk factors[88] and geographical spread[16]. In addition to identifying epidemiological associations with transmission, phylogenetic analyses have been used to identify anonymous contacts of rapidly progressive infections[198, 199], investigate outbreaks[178, 200], uncover outbreak sources[176, 201] and facilitate their reengagement into care[202] and guide further investigation to facilitate outbreak control[16, 176, 178]. Recently, in British Columbia, Canada, the use of near real-time phylogenetic surveillance of baseline HIV drug resistance tests was utilised as a public health tool, with individuals within a rapidly expanding phylogenetic cluster identified and targeted with interventions to reduce further transmission and attempt to identify further undiagnosed cluster members[189]. This work is the first to use routinely collected resistance test data in real-time to trace back to individuals with HIV, though similar approaches are being designed and used in the USA[203, 204]. Given the high coverage of resistance testing in the UK, real-time phylogenetic techniques may provide useful opportunities for identifying and curtailing outbreaks, such as that identified in Glasgow that had been spreading rapidly for some years prior its identification.

As there is limited experience and few published outcomes using real-time phylogenetics to guide interventions in HIV, it is unclear how it may be best utilised to aid case finding in the UK, and differences in patient management, criminalisation laws and partner notification policy may mean approaches need adapting. In addition, ethical and

acceptability issues must be considered[205, 206, 207]; using phylogenetic data to detect transmissions and trace contacts may raise a number of potential concerns and emotional reactions, including loss of privacy [208], guilt, blame [209], and prosecution[209, 210]. Although the use of phylogenetic data for public health intervention may be ethically justifiable if able to uncover undiagnosed infection and prevent onward spread, harm may result, for example reluctance to test, failure of disclosure of potentially infected contacts or refusal of resistance testing if patient and public factors are not adequately considered.

Given the limited published experience of using phylogenetics prospectively to guide public health interventions in HIV, this review aims to identify how molecular epidemiology has been used to focus case finding activities not only in HIV but also other similarly stigmatised infections in any population, to identify methods, ethical issues, barriers and facilitators have been identified within this context to help develop acceptable strategies in the UK.

2.3 Methods

2.3.1 Concepts and development of search terms (*see table 1*.)

In order to identify interventions relevant to the development of contact tracing for HIV, we sought to identify published studies according to the following three key dimensions;

Relevant interventions

We searched for studies, including outbreak reports where analysis of sequence data was directly used to focus contact tracing or testing activity for groups or individuals, using terms were included that described a potential mechanism of tracing. We focussed on case finding as linking contacts through phylogenetic networks may suggest a confirmation of transmission

Infections with comparable stigma

Though by nature, all infectious diseases may be stigmatised[211], we focussed on infectious diseases and modes of transmission widely considered to be associated with high levels

of stigma in the west, including Tuberculosis (TB), a potentially life threatening infection associated with high levels of stigma in the UK[212, 213], frequently associated with HIV infection, low income and migrant populations.

Molecular epidemiological methods

Terms were chosen to include only studies using phylogenetics, or molecular epidemiological methods as a basis for conducting potential tracing interventions.

Table 1: Search terms used for systematic review

Relevant intervention	Comparable stigma	Molecular epidemiological methods
Outbreak*	Sexually transmitted	Molecular epidemiology -
Outbreak - exp	disease – exp	exp
Contact tracing – exp	HIV*	"Molecular investigation*"
	"human immunodeficiency	"Molecular epidemiolog*"
(contact* or partner* or	virus"	Phylogen*
source* or link* or	Tuberculosis	Phylodynamic*
case*) adj9 (trac* or	TB	Phylogeograph*
notif* or investing* or	"Hepatitis B"	"Genetic evolution"
identif* or determin*)	"Hepatitis C"	sequenc*
	"Sexually transmitted"	genotyp*
(contact* OR partner*	"Sexual transmission"	
OR source* OR link*	gonorrhoea	
OR trac* OR notif* OR	chlamydia	
investing* OR identif*	Syphilis	
OR determin* OR case*)	Herpes	
adj9 (transmi*)		

Terms within a column combined with "OR", terms between columns combined with "AND".

2.3.2 Identification of papers

2.3.2.1 Database searches

On 27th May 2018, 4 Electronic databases (Medline, Embase, PsychInfo and CINAHL) were used to systematically identify papers using the search terms included in table 1. All identified references were exported into Zotero, an electronic reference management tool, and duplicates were removed.

2.3.2.2 Inclusion and exclusion criteria

We included papers which met all of the following criteria:

1. Any type of peer-reviewed study containing original data on human subjects
2. Studies that describe direct human-to-human transmission of a stigmatised/sexually transmitted infection
3. Studies that include the use of molecular epidemiology to directly guide or inform a contact tracing or testing strategy
4. Studies where tracing of contacts or targeted testing as a result of this data was planned, attempted or carried out.
5. Studies published in English.

Studies were excluded if they met any of the following criteria:

1. Studies describing nosocomial infection
2. Studies where molecular clusters of TB guided further investigation, including determining epidemiological linkages between diagnosed individuals, but did not report further active case finding of undiagnosed individuals being attempted.
3. Studies where contact tracing or testing procedures were independent of the molecular investigation.

Rationale for exclusion criteria

Nosocomial, including transfusion related transmission of HIV is now very rare, and policies and attitudes towards contacting potentially exposed individuals are likely to

differ greatly from contacting of sexual partners. As our study aims to identify methods, barriers and facilitators relevant to the tracing of community transmission of HIV, these studies were excluded.

Cluster investigation, analysing molecular clusters to identify epidemiological links between individuals is a standard surveillance method for investigating clusters of TB[214] and may be used to understand transmission dynamics in other infectious outbreaks. This may involve re-interview of patients to identify or confirm epidemiological links between cluster members to determine whether further investigation is needed. As this aims to inform ways of potentially uncovering undiagnosed HIV infection and barriers and facilitators to this, we excluded studies in which further active contact tracing of previously unidentified cases was not considered or required.

2.3.2.3 Selection of included papers

A title review was performed and non-relevant papers excluded. An abstract review was subsequently performed, and all papers meeting any exclusion criteria, or explicitly not meeting the inclusion criteria were excluded. Full text reviews of included articles were carried out, and eligible papers identified. In addition, a reference search was performed for all full texts reviewed. Any ambiguous papers were discussed with ***JC*** and a joint decision was made to exclude or include after review.

2.3.3 Quality assessment

Due to the heterogeneous nature of the included studies, standard quality assessment tools were not considered appropriate. The aims of this study were primarily to identify not only methods used to guide case finding, but ethical issues arising, facilitators and barriers, and standardised checklists may exclude papers containing relevant information. Therefore, to meet the objectives of this study, a quality assessment tool was developed to assess the clarity of the included papers in answering the review questions. This assessed;

1. The phylogenetic methods and the clarity of the rationale provided for their use.
2. Whether the role of phylogenetics in informing case finding was clear and whether this was a specific aim of the report

3. Whether case finding was carried out and if the methodologies used were clearly explained
4. Whether case finding outcomes were reported
5. Whether the methods used were directly translatable for use in HIV in the UK.

2.4 Results

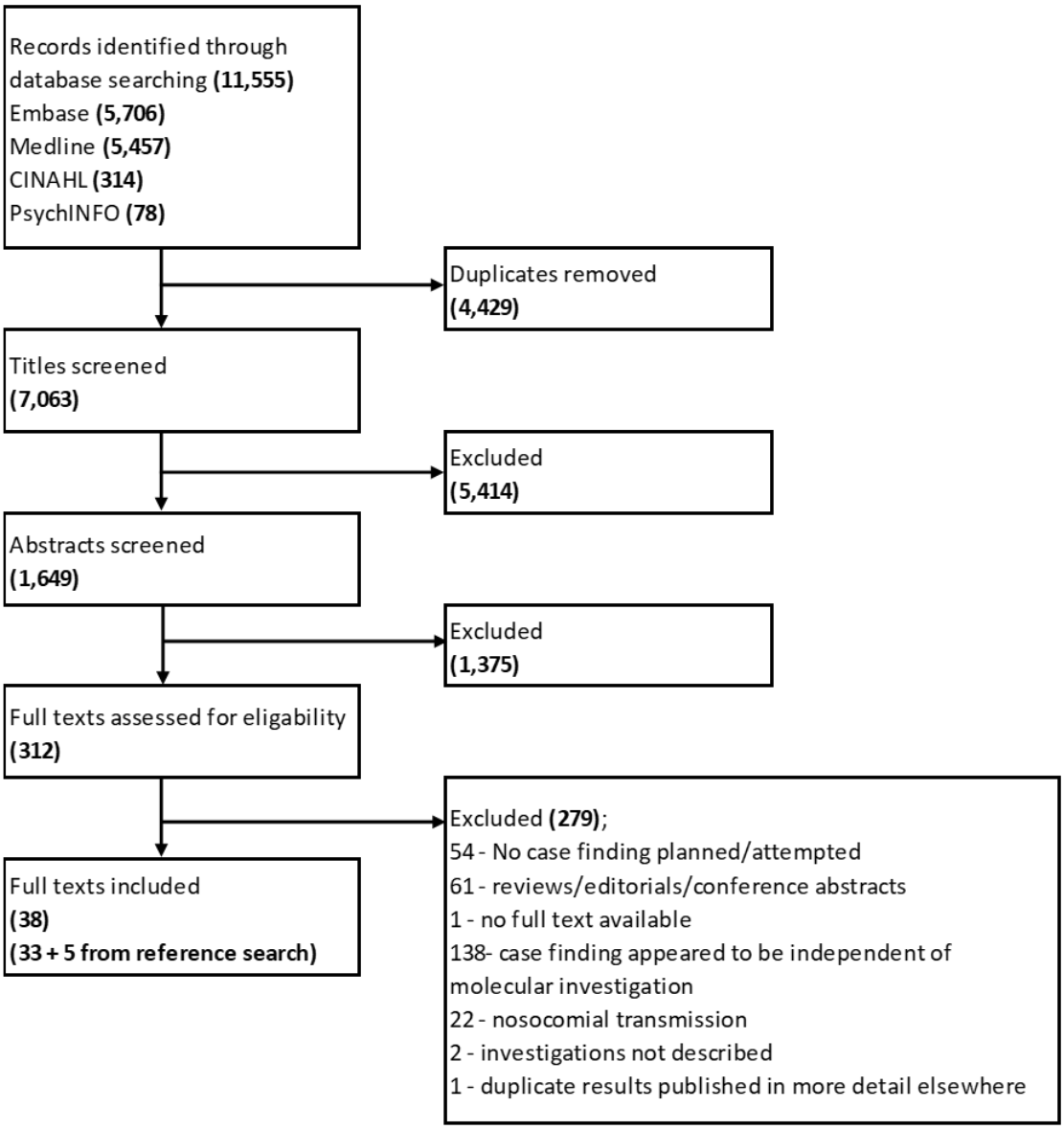


***Figure 11*: Prisma flowchart**

2.4.1 Study characteristics

38 studies were included in the review (Table 2). These included 31 studies of Tuberculosis and 7 studies of HIV infection. Three investigations were described over 2 papers (2 HIV, 1 TB), which were considered together, leaving 35 investigations. The studies were conducted between 1993 and 2017 and took place in the following countries; USA (n=18), The Netherlands (n=5), UK (n=4), Italy (n=3), Canada (n=1), Singapore (n=1), Greece (n=1), Australia (n=1) and 1 investigation took place in several European countries. The articles were heterogeneous in nature, consisting of outbreak reports, and implementation, observational and cross sectional studies. No studies describing case finding strategies based on phylogenetics of Hepatitis B or C, or other STIs were included, largely owing to the fact most published investigations described nosocomial transmission, or molecular epidemiology did not guide further case finding activities.

2.4.2 Quality of included papers

20 investigations were considered high quality in terms of addressing the study questions, with 15 considered as medium quality. For those not considered high quality, a variety of reasons were identified; case finding methods were not clearly explained in 6, the specific benefit of the addition of molecular epidemiology over traditional methods for case finding was not clear in 3 and reporting of contact tracing outcomes was unclear in 10. 30 did not appear to be directly translatable for the use of case finding for HIV in the UK, mainly as they focussed on a different infection, with different transmission dynamics. (Table 4).

2.4.3 Data synthesis

2.4.3.1 Methods identified

The methods of using genetic data to aid case finding fit into three themes. Some studies adopted more than one approach and are summarised below.

Identifying or confirming outbreak cases instigating further investigation

Most investigations (31) described the identification of outbreaks, or outbreak cases leading to further investigation. 28 studies investigated TB[215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242], 3 HIV[176, 177, 189, 243].

Several studies in TB reported the identification of previously unexpected clusters through routine surveillance or cluster review [216, 218, 220, 222, 223, 224, 226, 228, 229, 237, 240, 241], whilst in others, confirmation of recent transmission in an epidemiologically suspected cluster prompted further investigation[217, 219, 221, 225, 233, 234, 235, 242]. Case finding strategies included re-interviewing all clustered cases to identify previously unreported contacts, interviewing cluster members with no epidemiological linkages to determine common links for investigation, and identifying linked locations for further screening due to casual, unidentifiable contacts or poor recollection of names due to the retrospective nature of the interviews. Methods to assist in detecting unidentified contacts or sites of transmission included case notes review, interviews and cluster diagrams[223, 234], displaying both genetic and epidemiological linkages between cluster members to clarify transmission pathways.

We identified three studies where phylogenetic analysis was used to identify or confirm an outbreak in HIV, triggering case finding amongst potential contacts. In Doncaster, United Kingdom, phylogenetic analysis was used confirm a likely source in a suspected outbreak of HIV-1 infection[176]. Although not fully conclusive, the analysis was seen as sufficient evidence the outbreak was due to a single heterosexual source who had reported multiple episodes of unprotected intercourse with casual partners, prompting a media alert to encourage screening of potentially exposed contacts in the geographical region[177]. Similarly, in Seattle, USA, after the identification of several drug resistant HIV infections suspected to be an outbreak was confirmed to be related by phylogenetic analysis, cluster investigation with partner counselling and referral was performed for each clustered case to offer testing to all potentially exposed contacts[243]. Phylogenetic analysis was used to identify a previously unsuspected HIV outbreak during routine surveillance in British Colombia, Canada, where real-time phylogenetic analysis used as a identified a rapidly expanding cluster of drug resistant HIV infections[189]. Public Health officials were alerted, and interventions including contact tracing were employed to control the outbreak.

Epidemiological findings linked to molecular epidemiology used to guide the development of a targeted screening programmes

Three investigations led to the development of targeted screening programmes as a result of molecular epidemiological investigation. After a huge increase in newly diagnosed cases of HIV amongst people who inject drugs (PWID) in Athens, Greece, a phylogenetic analysis was performed to assess transmission networks and causal pathways of the outbreak[16]. The results, showing high levels of networking and a significant contribution of infections secondary to migration, particularly from Afghanistan and Iran, were used to guide a seek, test, treat and retain programme for PWID in Athens, with bilingual mediators guided by phylogeographic findings[244]. In Tarrant County, USA, geographical information systems analysis (GIS) was utilised to map zip-codes at the time of TB infection with TB strain typing[245]. This analysis identified specific locations of on-going transmission which were targeted for a successful location based screening programme[246].

Using linked epidemiological data from phylogenetic analysis to provide clues to a potential international transmission source

One investigation performed in Italy was identified in which a phylogenetic analysis of HIV was performed to investigate a case, diagnosed in Afghanistan, where routine contact tracing revealed no potential source of infection[247]. Analysis revealed the closest sequences to originate from Spain, which triggered in the index case a memory of a casual Spanish contact occurring shortly before a potential seroconversion illness. Unfortunately, the contact was found to be untraceable.

Table 2: Identified case finding themes

	Themes	Infections investigated	Studies
1	Cases with matching genotypes/clustered cases found leading to extended screening (i.e. detection of outbreak)	TB, HIV	Allen, KW. et al, 2002; Hayman, A. et al, 2001; Poon, AF. et al. 2016; Buskin, SE. et al, 2008; Faccini, M. et al.2013; Bifani, P.J et al. 1999; Merritt, TD et al. 2007; Lambregts-van Weezenbeek, CSB et al. 2003; Clark, CM et al. 2006; Ruddy, MD et al. 2004; Sebek, M et al. 2000; de Vries, G et al. 2009; Malakmadze N. et al, 2005; Kiers A et al, 1997; Black A et al, 2017; Ho ZJM et al, 2018; Lalor MK et al, 2018; Walker TM et al, 2018; Centers for Disease Control and Prevention (CDC), 2012; Centers for Disease Control and Prevention (CDC), 2005; Ashghar RJ et al, 2009; Bloss E et al, 2011; Lofy KH et al, 2006; Centers for Disease Control and Prevention (CDC)., 2009; Mitruka K et al, 2014; Miller AC et al, 2002; Samuel V et al, 2012; Buff AM et al, 2010
2	Characterisation of transmission sources within clusters leading to targeted screening of a specific population	TB, HIV	Moonan, PK et al. 2013; Moonan, PK. et al, 2004; Paraskevis, D et al. 2013; Hatzakis, A et al. 2015; de Vries G & van Hest RA, 2006
3	Epidemiological data retrieved from a phylogenetic connection providing support in identifying source	HIV	Ciccozzi, M. et al. 2011

3.4.3.2 Success of methods in identifying further infection

Success of interventions may be determined in different ways; firstly, though identification and treatment of undiagnosed individuals or evidence of slowing or resolution of an outbreak, or secondly through the lack of obstacles facilitating a complete investigation.

Only 18 papers reported total numbers screened/traced, and 19 reported the number of infected contacts (table 3.3, part 2). This may be due to the nature of the included studies, many reporting on technical aspects, or where the primary outcome of the study was not contact tracing. Additionally, there were no control groups reported in any of the studies. The majority of the included studies had reported, or suggested standard contact tracing had already occurred, meaning any additional contacts found could potentially be seen as 'success'. However, success in terms of the detection of new contacts guided by molecular epidemiology cannot be inferred from this review as investigations, including TB cluster investigations that may not have been taken forward due to futility or a lack of identifiable epidemiological links, were excluded, biasing these findings towards 'successful' contact tracing outcomes.

Two studies suggested interventions were successful by demonstrating resolution or slowing of an outbreak though phylogenetic analysis[179, 189], and others showed this by a reduction in identified outbreak cases over time, though again, in the absence of control groups this is not conclusive of benefit gained though molecular epidemiology. In the real-time investigation of HIV in Canada, further transmissions from non-drug resistant viruses occurred within a cluster where enhanced interventions were only applied to those with drug resistant viruses; this may reflect the success of the implemented interventions in the selected population, though may suggest these interventions should have been applied to the entire cluster[189].

In terms of barriers and facilitators affecting the success of investigations, studies reported various methodological and person related factors as discussed below.

2.4.3.3 Methodological and practical facilitators and barriers

Timing of investigations appeared to impact on the successes of the interventions, due to delays in sequencing, or delayed recognition of an outbreak. This led to poor recollection of epidemiological details from TB cluster cases in one investigation of international TB transmission[229], and in TB outbreak investigations in the USA, cluster members could not be re-interviewed as they had died, moved, or were at that point untraceable[224, 237, 238]. A further study reported difficulties collecting epidemiological data from TB cluster members due to their 'chaotic lifestyles'[228] and in the USA, UK and New South Wales

after investigation of clustered TB cases, newly identified locations of transmission had since closed and therefore could not be investigated further[217, 224, 226]. Timing also influenced investigation during TB 'cluster feedback' in the Netherlands – after interviewing clustered cases, many epidemiological links were from the distant past, or casual in nature, meaning re-opening contact tracing investigations was seldom thought to be worthwhile, with only 0.9% of clusters re-investigated[218]. Similarly in the investigation of an office TB outbreak in Italy, the retrospective nature of the investigation meant historic infections did not necessarily have genotypes, leading to incomplete reconstruction of transmission dynamics[241]. Rapid or 'real-time' sequencing and analysis was therefore found to facilitate investigations in outbreak situations[189, 219, 220], motivate investigators and improve co-operation with screening at homeless facilities by demonstrating the presence of outbreak strains of infection[234].

Sequencing methods appeared to facilitate two investigations using whole genome sequencing alongside standard methods for investigations of TB; WGS confirmed linkages that were considered to be due to chance similarity[227], and disproved previously identified genetic clusters, allowing resources to be focussed elsewhere[228].

Cost and human resource was identified as a barrier to the potential use of investigations of TB guided by molecular epidemiology [218, 221, 226, 228, 237, 248]. In one study, the additional costs required to perform extended screening of a TB outbreak were in part covered by a research grant, which would not be available in non-research situations[221], though government strategies for managing outbreak investigations, including the adequate provision of resources were implemented shortly after this incident. Significant costs and expertise required were reported in Universal TB genotyping in New York City, Massachusetts and the Netherlands[218, 220], though this was balanced by the benefits seen, and it was proposed that national implementation in the USA would reduce costs locally. Poon et al similarly argue for cost efficiency in real-time phylogenetic analysis of HIV, given the data utilised is already collected and processed as part of routine care[189].

In addition, *anonymous or casual sex* was a barrier to successful contact tracing in studies of HIV infection[177, 243, 247], as were casual or unknown contacts in TB[218] though this is a universal issue unrelated to molecular methods used. Clustering in fact facilitated the identification of casual/unknown contacts in Tuberculosis by linking locations of

clustered individuals, and helped identify cluster cases in situations where reporting of names was poor.

2.4.3.4 Person-related facilitators and barriers

Few patient barriers or facilitators were directly addressed. Two studies (TB and HIV) reported refusal of clustered individuals to partake in repeat contact tracing interviews, though the reasons for this were not stated[189, 237]. Several studies of TB reported low numbers of named contacts, which may be due to recollection or issues with disclosure; one study reporting reluctance for patients to provide locations of linked 'crack houses', identified as potential TB transmission sites[224], and five reporting a reluctance to name contacts or disclose epidemiological data, the authors hypothesising this may be due to stigma, involvement in illicit behaviours or fear of data being passed on to immigration authorities[226, 229, 232, 235, 240]. Screening of 'observed' (people attending linked venues) rather than named contacts was consequently shown to provide a higher yield of new latent TB contacts in one study[232]. Potential issues around stigma and disclosure were also addressed in an outbreak of TB in an apartment block with the inclusion of alternate site screening locations[227].

One identified barrier was the refusal by a landlord to allow screening of their venue, identified as a site of TB transmission[223], however informing staff of the presence of a TB outbreak strain amongst residents improved co-operation with investigation at a homeless shelter[234]. Community engagement was reported as a facilitator in two further investigations of TB. When using geographical information systems to guide location based screening of TB, there was initial reluctance from community organisations to engage with proposed screening interventions[246]. However, after presented the findings of the cluster analysis, revealing the potential positive impact to the community of the intended project, organisational support was obtained and a successful screening intervention implemented. Similarly, the assistance of the media and grassroots volunteers facilitated mass screening after an outbreak in a high-rise apartment block[227].

2.4.3.5 Ethical governance and issues raised

As most studies reported public health investigations, only five had documented a statement of ethical approval, with seven further studies explicitly stating why this was not required. In one TB outbreak investigation involving several countries, procedures were carried out under public health laws in all countries apart from Switzerland, where ethical approval was required.

The methods used to maintain patient confidentiality were discussed in eight papers. These included; the removal of patient identifiers prior to genetic analysis[16, 189, 221, 222]; only releasing identities to contacts with consent[217] or avoiding the use of contact names or pictures to trigger memory[240]; only investigating clusters comprising a minimum of 5 individuals to minimise attribution of transmission to any individual[189]; ensuring minimal data required for public health investigation was sent in a secure encrypted format[189]; providing alternate locations for screening in the investigation of an apartment block[227]; and withholding the release of a sources identity to the media in a heterosexual outbreak of HIV, despite potential benefits for case finding[177].

In a Canadian HIV outbreak, the balance between individual's autonomy and public health needs was debated in relation to the use of real-time HIV phylogenetics, authors concluding that the need to prevent HIV transmission was priority[189]. These investigations were performed under the authority of the Public Health Act. HIV is a reportable disease in Canada, therefore the authorities to whom the clustered individual's information was shared were already aware of their identities, with safeguards to keep data secure when being transferred. Irrespectively, careful consideration was given to the balance between maintaining confidentiality and preventing further HIV infections, with multi-disciplinary team discussion regarding clusters prior to the initiation of further investigation.

Although not explicitly mentioned, this issue of balance between autonomy and public health is intertwined with the concept of an outbreak investigation as a whole, as reflected by the low rates of ethical approval required within these studies. Most of the included studies were carried out under the ethical umbrella of 'public health and surveillance' as urgent action was required, meaning patient consent to use their infection's genetic data

Table 3: Study Characteristics, part 1

*Author, date of publication**	*Infection*	*Type of study*	*Year of investigation*	*Location*	*Theme***	*Aims*	*Population*	*Genotyping/ molecular methods*	*Definition of clustering*	*Case finding approach*
Allen, KW. et al, 2002; Hayman, A. et al, 2001	HIV	Outbreak investigation	1998	Doncaster, UK	1	Determine the source of an outbreak; Describe the resultant public health investigation	Presumed outbreak cases (n=13) and potential contacts within Doncaster	Neighbour-joining and maximum likelihood trees of full length gp120 sequences	Bootstrap support (98%)	Public health investigation and televised alert to prompt testing of potential contacts guided by identification of source
Poon, AF. et al. 2016	HIV	Implementation case study of outbreak investigation	2014	British Colombia, Canada	1	Describe a surveillance intervention	All sequenced HIV positive individuals in British Colombia (n= 8839). Contact investigation of new clustered cases with resistant HIV (n=9)	Maximum likelihood trees of *pol* gene	>/=5 sequences, patristic distance < 0.02 , >/= 50% bootstrap support	Individuals within a rapidly growing cluster underwent repeat contact tracing
Buskin, SE.et al, 2008	HIV	Outbreak investigation	2005-2007	King County, Seattle, USA	1	Describe an outbreak investigation	All local newly-diagnosed patients with multi-drug resistant HIV (n = 9)	Neighbour-joining and maximum likelihood trees of *pol* sequences	Compared genetic distance with random controls to demonstrate significance	Contact tracing of phylogenetically confirmed outbreak cases

*Author, date of publication**	*Infection*	*Type of study*	*Year of investigation*	*Location*	*Theme**	*Aims*	*Population*	*Genotyping/ molecular methods*	*Definition of clustering*	*Case finding approach*
Paraskevis, D et al. 2013; Hatzakis, A et al. 2015	HIV	Observational - Longitudinal phylogenetic and phylogeographic study and descriptive study of targeted outreach programme	2008-2012	Athens, Greece	2	Describe an outbreak; Describe and assess an intervention implemented in response to the outbreak	All HIV positive intra-venous drug users in Athens with a HIV sequence collected from 1998 - 2012 (n=282)	Maximum likelihood and Bayesian analysis of *pol* sequences	Shimodaira-Hasegawa values >0.95	Targeted outreach screening programme guided by findings from phylogenetic analysis
Ciccozzi, M. et al. 2011	HIV	Case report	2009	Italy	3	Describe a source finding exercise	1 individual with international sequences	Neighbour-joining and maximum likelihood trees of *pol* sequences	Bootstrap support >70%	Epidemiological data linked to sequences used in phylogenetic analysis helped prompt patient for source identification
Faccini, M. et al.2013	TB	Outbreak investigation	2010-2011	Milan, Italy	1	To describe an outbreak investigation	Local school and homeless population	Not described	Not defined	Detected epidemiologically unrelated clusters leading to outbreak investigation
Bifani, P.J et al. 1999	TB	Population based cross sectional study	1996 - 1998	New Jersey, USA	1	To investigate the spread of a resistant strain of TB	All local genotyped W family TB cases (n > 68)	Analysis of sequences derived by IS6110 DNA fingerprinting, PGRS typing, spoligotyping & VNTR	Matching strains	Reopened contact tracing investigations within a cluster

*Author, date of publication**	*Infection*	*Type of study*	*Year of investigation*	*Location*	*Theme**	*Aims*	*Population*	*Genotyping/ molecular methods*	*Definition of clustering*	*Case finding approach*
Merritt, TD et al. 2007	TB	Outbreak investigation	1994 - 2005	Hunter Area, New South Wales, Australia	1	Characterise a cluster of TB; Describe the public health response	All genotyped TB cases reported in the Hunter area between January 1994 and June 2005 (n=21) and epidemiologically linked non-genotyped cases (n=2)	Analysis of sequences derived by MIRU spoligotyping followed by IS1660 RFLP analysis for matching types	Cases indistinguishable by MIRU, spoligotyping and IS6110 RFLP analysis, or those with only 1 missing/different band in RFLP if epidemiological link confirmed	Repeat interviews for clustered individuals to detect undiagnosed contacts and identify linked venues for screening
Moonan, PK. et al, 2004; Moonan, PK. et al, 2006	TB	Observational/ implementation study	1993 - 2000, 2002 - 2004	Tarrant County, USA	2	Identify geographical areas of on-going TB transmission; Evaluate genotyping-GIS screening of TB	All local genotyped TB cases (527). Screening for all individuals from high prevalence zip codes accessing community organisations (3,645)	Analysis of sequences derived by IS6110-based RFLP and spoligotyping	Identical IS6110 (>7 bands) or identical IS6110 & spoligotype (<7 bands)	Geographically targeted screening based on cluster analysis
Lambregts-van Weezenbeek, CSB et al. 2003	TB	Implementation study	1995-2000	The Netherlands	1	Describe a DNA surveillance programme and its use in contact investigation	All clustered TB cases in the Netherlands(n= 3854) plus contacts	Analysis of sequences derived by IS6110 RFLP and PGRS	Matching IS6110 if >4 bands, IS6110 & matching PGRS if <5	Reopening of contact investigations after 'cluster feedback'

*Author, date of publication**	*Infection*	*Type of study*	*Year of investigation*	*Location*	*Theme**	*Aims*	*Population*	*Genotyping/ molecular methods*	*Definition of clustering*	*Case finding approach*
Ashworth, M et al. 2008	TB	Outbreak investigation	2004 - 2006	King County, Washington, USA	1	Evaluate two PCR based genotyping approaches; Describe an outbreak investigation	Epidemiologically linked presumed outbreak cases (n=5) plus contacts	Analysis of sequences derived by 12-locus MIRU and rep-PRC	>92% similarity by rep-PCR	Re-interview of clustered cases. Further contact tracing of unexpected genotype matches
Clark, CM et al. 2006	TB	Implementation study	2001-2003	New York City, USA	1	Describe the use of routine genotyping in determining transmission dynamics and refining public health investigations	All genotyped TB cases in New York City (n=2, 437) plus potential contacts	Analysis of sequences derived by IS6110 RFLP and spoligotyping	>1 sequence with identical IS6110 RFLP banding pattern and spoligotype	Expanded testing in locations associated with clustering
Ruddy, MD et al. 2004	TB	Outbreak investigation	2000-2001	London, UK	1	Describe an outbreak investigation	Presumed/confirmed outbreak cases (n= 93) and contacts	Analysis of sequences derived by IS6110 RLFP +/- RAPET	Identical 15 band RFLP-IS6110 fingerprint and RAPET profile	Contact tracing of clustered cases, expanded screening of locations associated with clustering
Sebek, M et al. 2000	TB	Implementation study	1995-1997	The Netherlands	1	Describe a DNA fingerprinting surveillance project and its use in contact tracing	Molecularly clustered TB cases in the Netherlands (n=2,218)	Analysis of sequences derived by IS6110 RFLP analysis +/- PGRS	>1 100% identical strains	Re-interviewed clustered patients and reopened contact tracing investigations within unexpected clusters

*Author, date of publication**	*Infection*	*Type of study*	*Year of investigation*	*Location*	*Theme***	*Aims*	*Population*	*Genotyping/ molecular methods*	*Definition of clustering*	*Case finding approach*
de Vries, G et al. 2009	TB	Outbreak investigation	2002 - 2007	Rotterdam, The Netherlands	1	Describe the use of a novel cluster diagram to guide public health interventions	Molecularly clustered cases and epidemiologically linked culture negative cases (n=32)	Analysis of sequences derived by IS6110 RFLP	Identical RFLP or by Polymorphic GC-rich sequence probe if <5 IS1660 copies	Expanded screening of locations associated with clustering
Malakmadze N. et al, 2005	TB	Outbreak investigation	2003	Wisconsin, USA	1	Describe an outbreak investigation	Individuals within genotype clusters (n = 19) plus contacts	Analysis of sequences derived by MIRU and IS6110 RLFP	Not defined	Investigation of clustered individuals with few epidemiological links to determine undiagnosed contacts and locations of transmission for screening
Kiers A et al, 1997	TB	Outbreak investigation	1993-1996	The Netherlands	1	Describe an outbreak investigation		Analysis of sequences derived by RLFP	Matching strain types	Investigation of non-epidemiologically linked clustered cases led to identification of source and screening of locations associated with transmission
Black A et al, 2017	TB	Outbreak Investigation	2000	Southampton, UK	1	Describe an outbreak investigation	TB cases in Southampton	Analysis of sequences derived by 24-locus MIRU-VNTR and WGS	Indistinguishable genotypes	repeat interviews and questionnaires led to screening of contacts and linked locations names by cluster members

*Author, date of publication**	*Infection*	*Type of study*	*Year of investigation*	*Location*	*Theme**	*Aims*	*Population*	*Genotyping/ molecular methods*	*Definition of clustering*	*Case finding approach*
Ho ZJM et al, 2018	TB	Outbreak Investigation	2012-2016	Singapore	1	Describe an outbreak investigation	Outbreak cases in Singapore	Analysis of sequences derived by 24-locus MIRU-VNTR, Spoligotype and WGS	Identical MIRU-VNTR and Spoligotype	Identification of an outbreak with few epidemiological links led to mass screening of a high-rise apartment block
Lalor MK et al, 2018	TB	Outbreak Investigation	2012	UK	1	Describe an outbreak investigation and the impact of WGS	18 clustered outbreak cases	Analysis of sequences derived by 24-locus MIRU VNTR and WGS, ML phylogenetic trees	<12 SNP difference between isolates	Cluster investigation led to identification of two new settings for public health action
Walker TM et al, 2018	TB	Outbreak Investigation	2016-2017	Europe	1	Describe an outbreak investigation	29 outbreak cases identified across Europe	Analysis of sequences derived by 24-locus MIRU VNTR and WGS	<5 SNP difference between isolates	Interviews of clustered outbreak cases to identify routes and shared locations of transmission and perform contact investigation
Centers for Disease Control and Prevention (CDC), 2012	TB	Outbreak Investigation	2010-2011	Illinois, USA	1	Describe an outbreak investigation	6 outbreak cases linked to an overnight shelter	Analysis of sequences derived by spoligotype & 12-locus MIRU-VNTR	Matching genotypes	Identification of outbreak led to on site contact tracing and mass screening

*Author, date of publication**	*Infection*	*Type of study*	*Year of investigation*	*Location*	*Theme**	*Aims*	*Population*	*Genotyping/ molecular methods*	*Definition of clustering*	*Case finding approach*
Centers for Disease Control and Prevention (CDC), 2005	TB	Outbreak Investigation	2003	New York City, USA	1	Describe an outbreak investigation	4 clustered TB cases linked to a homeless shelter	Analysis of sequences derived by IS6110-RFLP and spoligotype		Identification of an outbreak led to screening of staff and shelter residents
Ashghar RJ et al, 2009	TB	Outbreak Investigation	2004-2005	Miami, USA	1	Describe an outbreak investigation	18 clustered TB cases linked by crack cocaine use in a Miami neighbour-hood	Analysis of sequences derived by MIRU , IS6110-RFLP & spoligotype	matching spoligotypes	Clustered cases re-interviewed for contacts and locations for screening
Bloss E et al, 2011	TB	Outbreak Investigation	2007	Jackson County, USA	1	Describe dynamics of transmission and outbreak response	11 clustered TB cases in the study region	Analysis of sequences derived by spoligotyping, MIRU-VNTR	matching genotypes or social links	Interview of clustered cases to identify contacts
Lofy KH et al, 2006	TB	Outbreak Investigation	2002-2003	Washington, USA	1	Describe an outbreak investigation		Analysis of sequences derived by spoligotyping, MIRU & IS6110-RFLP	Matching genotypes	Confirmation of a suspected outbreak led to intensified investigation and screening of a shelter

*Author, date of publication**	*Infection*	*Type of study*	*Year of investigation*	*Location*	*Theme**	*Aims*	*Population*	*Genotyping/ molecular methods*	*Definition of clustering*	*Case finding approach*
Centers for Disease Control and Prevention (CDC), 2009	TB	Outbreak Investigation	2004-2007	Michigan, USA	1	Describe an outbreak investigation	8 genotypically clusterd TB cases in Detroit	Analysis of sequences derived by spoligotype, MIRU and RFLP	Matching spoligotype and MIRU patterns	Identification of outbreak led to intensified contact tracing
Mitruka K et al, 2014	TB	Outbreak Investigation	2006-2008	Nevada and Arizona, USA	1	Describe an outbreak investigation	8 clustered TB cases in a local Hispanic community	Analysis of sequences derived by spoligotype and 12-locus MIRU	Matching genotypes	Identification of an outbreak led to re-interview of clustered cases, review of cell-phone directories and interview of contacts if named by multiple cases to identify further contacts
Miller AC et al, 2002	TB	Implementation study	1996-2000	Massachusetts, USA	1	Describe the impact of TB genotype surveillance on public health practice	Analysis of sequences derived by IS6110 RFLP & Spoligotyping	Identical IS6110 RFLP patterns in high copy strains, Identical RFLP & spoligotype in low copy strains	State-wide genotype cluster review with identification of unexpected epidemiological links. Identification of cluster led to mass screening of a homeless shelter
Samuel V et al, 2012	TB	Outbreak Investigation	2012	Florida, USA	1	Describe an outbreak investigation	Analysis of sequences derived by spoligotyping & 12-locus MIRU-VNTR	Matching genotypes	An increased proportion of a specific genotype amongst a homeless population led to mass screening

*Author, date of publication**	*Infection*	*Type of study*	*Year of investigation*	*Location*	*Theme**	*Aims*	*Population*	*Genotyping/ molecular methods*	*Definition of clustering*	*Case finding approach*
Buff AM et al, 2010	TB	Outbreak Investigation	2005-2007	South Carolina, USA	1	Describe an outbreak investigation	21 clustered TB cases and contacts	Analysis of sequences derived by spoligotyping & 12-locus MIRU-VNTR	Matching genotypes	All clusters within the state genotyping surveillance system were ranked and the largest investigated. When new cases of this genotype are detected, site based screening is initiated.
de Vries G & van Hest RA, 2006	TB	Outbreak Investigation	2001	Rotterdam, The Netherlands	2	Describe an outbreak investigation	507 drug users, homeless individuals and staff at shelters	Analysis of sequences derived by RFLP	Matching RFLP pattern	Reintroduction of targeted screening on finding different genotypes amongst an outbreak
McConkey S. et al, 2002	TB	Outbreak Investigation	1998-1999	Missouri, USA	1	Evaluate molecular typing as a method for TB control	62 culture confirmed TB cases in the study area	Analysis of sequences derived by IS6110 RFLP, PGRS	Identical IS6110 with 15 hybridizing bands or 15 hybridizing bands differing by a single band with an identical PGRS; or identical IS6110 with <6 bands and an identical PGRS	Identification of clustered individuals led to repeat contact tracing
Faccini, M. et al.2015	TB	Outbreak Investigation	2011	Milan, Italy	1	Describe an outbreak investigation	TB cases and contacts within an office	Analysis of sequences derived by 24-locus MIRU-VNTR		Identification of 2 genotypically and epidemiologically linked cases diagnosed at different times led to expanded testing of workplace

*Author, date of publication**	*Infection*	*Type of study*	*Year of investigation*	*Location*	*Theme* **	*Aims*	*Population*	*Genotyping/ molecular methods*	*Definition of clustering*	*Case finding approach*
Powell, KM. et al, 2017	TB	Outbreak Investigation	2008-2015	Georgia, USA	1	Describe an outbreak investigation	110 TB outbreak cases in Georgia	Analysis of sequences derived by spoligotyping and 24-locus MIRU-VNTR		Re-interview of clustered cases and screenings of shelters that had homed outbreak cases

Abbreviations: PGRS = polymorphic GC-rich tandem repeat; VNTR = variable numbers of tandem repeats; MIRU = mycobacterial interspersed repetitive unit typing; RFLP = restriction fragment length polymorphism; rep-PCR = repetitive-unit-sequence-based PCR; RAPET = rapid epidemiological typing; WGS = whole genome sequencing; SNP = single-nucleotide polymorphism

* Linked studies analysed together

** 1. Identifying or confirming outbreak cases instigating further investigation 2. Characterisation of transmission sources within clusters leading to targeted screening of a specific population. 3. Epidemiological data retrieved from phylogenetic connection providing support in identifying source.

Table 3: Study characteristics, part 2

Author, date of publication	Further potential contacts identified? (Number if reported)	Further contacts tested? (Number if reported)	New infections detected? (Number if reported)	Rate of positive testing in contacts (numbers positive/ screened)	Reported barriers	Reported facilitators	Ethical issues addressed	Methods to maintain confidentiality	Ethics statement present
Allen, KW. et al, 2002 Hayman, A. et al, 2001	Yes	Yes (772)	Yes (2)	0.30%			Debate on revealing sources identity to the press	Investigators kept sources identity confidential and provided safe accommodation	No
Poon, AF. et al. 2016	Yes (12)	Yes (5)	Yes (1)	20%	2 of 9 individuals declined repeat contact tracing - no explanation why	MDT discussion of clusters. Sequences derived from leftover serum of every first viral load, increasing sampling density	Balance between individual autonomy and public health needs. Care taken to ensure confidentiality and avoid blame. Consent received for contact tracing. Legal issues considered.	Patient identifiers replaced prior to phylogenetic analysis; minimum patient data sent in secure encrypted format to public health authorities	No***

Author, date of publication	Further potential contacts identified? (Number if reported)	Further contacts tested? (Number if reported)	New infections detected? (Number if reported)	Rate of positive testing in contacts (numbers positive/ screened)	Reported barriers	Reported facilitators	Ethical issues addressed	Methods to maintain confidentiality	Ethics statement present
Buskin, SE.et al, 2008	Yes	Yes (<45)	Yes (Not quantified)		High rates of anonymous sex made tracing difficult but linked individuals to venues		Discuss lack of need for ethical approval as investigations conducted as part of a surveillance programme		No***
Paraskevis, D et al. 2013	Yes	Yes (3320)	Yes (499)	15.00%				Phylogenetic and epidemiological data linked and anonymised prior to analysis	Yes
Hatzakis, A et al. 2015									
Ciccozzi, M. et al. 2011	Yes (1)	No (0)	No	0					No
Faccini, M. et al.2013	Yes	Yes (977)	Yes (15 TB, 173 LTB)	19%					No

Author, date of publication	Further potential contacts identified? (Number if reported)	Further contacts tested? (Number if reported)	New infections detected? (Number if reported)	Rate of positive testing in contacts (numbers positive/ screened)	Reported barriers	Reported facilitators	Ethical issues addressed	Methods to maintain confidentiality	Ethics statement present
Bifani, P.J et al. 1999	Yes	Yes	Yes (5)						No
Merritt, TD et al. 2007	Yes	Yes	No	0%	Retrospective nature of investigation hindered identification of contacts		No but brought to light the association with a music studio patients had not initially wished to disclose	Patient consent gained prior to releasing names to other cluster members	No
Moonan, PK. et al, 2004; Moonan, PK. et al, 2006	Yes	Yes	Yes (44 TB, 681 LTB)	19.9% (1.2% TB, 18.6% LTB)		Community support gained by sharing data revealing potential public health impact	Community consent gained for screening		Yes

Author, date of publication	Further potential contacts identified? (Number if reported)	Further contacts tested? (Number if reported)	New infections detected? (Number if reported)	Rate of positive testing in contacts (numbers positive/ screened)	Reported barriers	Reported facilitators	Ethical issues addressed	Methods to maintain confidentiality	Ethics statement present
Lambregts-van Weezenbeek, CSB et al. 2003	Yes	Yes	Yes (12 TB, 71 LTB)						No
Ashworth, M et al. 2008	Not reported	Not reported	Not reported			Real-time sequencing and investigation			Yes
Clark, CM et al. 2006	Yes	Yes	Yes (4 TB, additional LTB)			Real-time investigation assisted by co-operation and timely submission by participating laboratories			No***
Ruddy, MD et al. 2004	Yes	Yes (269, screening ongoing at time of publication)	Yes (26)	11%	Cost and human resource required			Data de-identified prior to analysis	No

Author, date of publication	Further potential contacts identified? (Number if reported)	Further contacts tested? (Number if reported)	New infections detected? (Number if reported)	Rate of positive testing in contacts (numbers positive/ screened)	Reported barriers	Reported facilitators	Ethical issues addressed	Methods to maintain confidentiality	Ethics statement present
Sebek, M et al. 2000	Yes	Yes (1904)	Yes (48 LTB, 6 TB)	2.80%			Patient privacy maintained	Names only released directly to TB nurse, with database data anonymised	No
de Vries, G et al. 2009	Yes	Yes (2328)	Yes (6 TB, 75 LTB)	5.80%	Landlord refused to co-operate with screening of venue	Cluster diagram aided visualisation of linked locations			Yes
Malakmadze N. et al, 2005	Yes (98)	Yes (30)	Yes (5 LTB)	17%	Participants refused to provide the locations of 'crack houses'				No
Kiers A et al, 1997	Yes	Yes (>1000)	Yes (>24)						No

Author, date of publication	Further potential contacts identified? (Number if reported)	Further contacts tested? (Number if reported)	New infections detected? (Number if reported)	Rate of positive testing in contacts (numbers positive/ screened)	Reported barriers	Reported facilitators	Ethical issues addressed	Methods to maintain confidentiality	Ethics statement present
Black A et al, 2017	Yes (98)	Yes (74)	Yes (7TB, 24 LTB)		Reluctance to name contacts; Resource intensive; retrospective nature led to incomplete investigation	Public health awareness raising			
Ho ZJM et al, 2018	Yes (373)	Yes (259)	Yes (5TB, 46 LTB)			Assistance of media and grassroots volunteers, WGS confirmed linkages that may have been presumed as chance MIRU links		alternate Alternate screening locations provided due to stigma	
Lalor MK et al, 2018	Yes	Yes	No		Resource intensive, difficulty in collecting epidemiological data due to 'chaotic lifestyles' of outbreak members	WGS disproved clusters, allowing focussing of resources			No***
Walker TM et al, 2018	Not reported	Not reported	Not reported		Retrospective investigation - poor memory of exposure sites; suspected fear of disclosure of information to immigration authorities				Yes

Author, date of publication	Further potential contacts identified? (Number if reported)	Further contacts tested? (Number if reported)	New infections detected? (Number if reported)	Rate of positive testing in contacts (numbers positive/ screened)	Reported barriers	Reported facilitators	Ethical issues addressed	Methods to maintain confidentiality	Ethics statement present
Centers for Disease Control and Prevention (CDC), 2012	Yes	Yes (386)	Yes (8 TB, 146 LTB)						No
Centers for Disease Control and Prevention (CDC), 2005	Yes (1,335)	Yes (1,186)	Yes (4 TB, 239 LTB)						No
Ashghar RJ et al, 2009	Yes	Yes (187)	Yes (21 LTB)		Reluctance of cases to name names	Screening of 'observed' contacts attending reported locations, due to difficulty eliciting names			No
Bloss E et al, 2011	Yes (170)	Yes (157)	Yes (53 LTB)						No***

Author, date of publication	Further potential contacts identified? (Number if reported)	Further contacts tested? (Number if reported)	New infections detected? (Number if reported)	Rate of positive testing in contacts (numbers positive/ screened)	Reported barriers	Reported facilitators	Ethical issues addressed	Methods to maintain confidentiality	Ethics statement present
Lofy KH et al, 2006	Yes	Yes (1000)	Yes (13 TB, 111 LTB		Poor reporting of named contacts	Site based screening in the absence of contact names; Rapid genotyping feedback motivated investigators and improved cooperation at homeless facilities by verifying outbreak strain.			No ***
Centers for Disease Control and Prevention (CDC), 2009	Yes (79)	Yes (51)	Yes (5 LTB)		Refusal of index case to name contacts				No
Mitruka K et al, 2014	Yes (667)	Yes (590)	Yes (130 LTB)		Drug use and fear of deportation limited completeness of contact investigation	Genotyping linked cases that were not named due to stigma			No***

Author, date of publication	Further potential contacts identified? (Number if reported)	Further contacts tested? (Number if reported)	New infections detected? (Number if reported)	Rate of positive testing in contacts (numbers positive/ screened)	Reported barriers	Reported facilitators	Ethical issues addressed	Methods to maintain confidentiality	Ethics statement present
Miller AC et al, 2002	Yes	Yes	No		Sequencing expensive; microbiological expertise required; lengthy turnaround; difficult to reopen cases when time elapsed; patients unreachable or refused repeat interviews; untypeable specimens	Cluster review identified unknown sites, unexpected clusters and disproved outbreaks saving unnecessary investigation and resources			Yes
Samuel V et al, 2012	Yes (4,400)	Yes (2,300)	Not reported		Retrospective investigation – 13 clustered cases had dies by the time the investigation took place				No
Buff AM et al, 2010	Yes	Yes	Not reported				Informed consent gained from investigated cluster members		No***

Author, date of publication	Further potential contacts identified? (Number if reported)	Further contacts tested? (Number if reported)	New infections detected? (Number if reported)	Rate of positive testing in contacts (numbers positive/ screened)	Reported barriers	Reported facilitators	Ethical issues addressed	Methods to maintain confidentiality	Ethics statement present
de Vries G & van Hest RA, 2006	Yes	Yes	Not reported						No
McConkey S. et al, 2002	No	0	0		Patients who were homeless or drug users named few places and were reluctant to name contacts involved in illicit behaviours.				Yes
Faccini, M. et al.2015	Yes	Yes (107)	Yes (30 LTB)		Retrospective nature led to incomplete reconstruction of transmission dynamics; TB related stigma led to incomplete contact tracing and subsequent outbreak				No

Author, date of publication	Further potential contacts identified? (Number if reported)	Further contacts tested? (Number if reported)	New infections detected? (Number if reported)	Rate of positive testing in contacts (numbers positive/ screened)	Reported barriers	Reported facilitators	Ethical issues addressed	Methods to maintain confidentiality	Ethics statement present
Powell, KM. et al, 2017	Yes	Yes	Yes (2)		Limited utility of name based contact tracing; poorly kept rosters at homeless shelters; unsuccessful genotyping in some culture confirmed cases				No***

*** Ethical approval considered in manuscript but not required LTB= latent tuberculosis infection

Table 4: Quality assessment

Study	Were the phylogenetic/ molecular methods clearly described?	Was the rationale behind the use of phylogenetics clear?	Was case finding a specific aim?	Were the methods used for case finding clearly explained?	Was it clear how phylogenetics aided case finding?	Was case finding performed?	Were undiagnosed cases found?	Directly translatable for use with HIV in the UK?	Quality*
Allen, KW et al, 2002, Hayman, A et al, 2001	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	High
Poon, AF et al. 2016	Yes	Yes	?	Yes	Yes	Yes	Yes	Yes	High
Buskin, SE.et al, 2008	Yes	Yes	Yes	Yes	?	Yes	No	Yes	Med
Paraskevis, D et al. 2013, Hatzakis, A et al. 2015	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	High
Ciccozzi, M. et al. 2011	Yes	Yes	Yes	N/A	Yes	No	No	Yes	Med
Faccini, M. et al.2013	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	High
Bifani, P.J et al. 1999	Yes	Yes	No	No	No	Yes	Yes	No	Med
Merritt, TD et al. 2007	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Med
Lambregts-van Weezenbeek, CSB et al.	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Med
Ashworth, M et al. 2008	Yes	Yes	No	No	Yes	Yes	?	No	Med
Clark, CM et al. 2006	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	High

Study	Were the phylogenetic/ molecular methods clearly described?	Was the rationale behind the use of phylogenetics clear?	Was case finding a specific aim?	Were the methods used for case finding clearly explained?	Was it clear how phylogenetics aided case finding?	Was case finding performed?	Were undiagnosed cases found?	Directly translatable for use with HIV in the UK?	Quality*
Ruddy, MD et al. 2004	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	High
Sebek, M et al. 2000	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	High
de Vries, G et al. 2009	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	High
Malakmadze N. et al, 2005	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	High
Kiers A et al, 1997	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	High
Moonan, PK. et al, 2004; Moonan, PK. et al, 2006	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	High
Black A et al, 2017	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	High
Ho ZJM et al, 2018	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	High
Lalor MK et al, 2018	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Med
Walker TM et al, 2018	Yes	Yes	Yes	No	Yes	?	?	No	Med
Centers for Disease Control and Prevention (CDC), 2012	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	High

Study	Were the phylogenetic/ molecular methods clearly described?	Was the rationale behind the use of phylogenetics clear?	Was case finding a specific aim?	Were the methods used for case finding clearly explained?	Was it clear how phylogenetics aided case finding?	Was case finding performed?	Were undiagnosed cases found?	Directly translatable for use with HIV in the UK?	Quality*
Centers for Disease Control and Prevention (CDC), 2005	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	High
Ashghar RJ et al, 2009	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	High
Bloss E et al, 2011	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	High
Lofy KH et al, 2006	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	High
Centers for Disease Control and Prevention (CDC)., 2009	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Med
Mitruka K et al, 2014	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	High
Miller AC et al, 2002	Yes	Yes	Yes	No	Yes	Yes	?	No	Med
Samuel V et al, 2012	Yes	Yes	Yes	Yes	Yes	Yes	?	No	Med
Buff AM et al, 2010	Yes	Yes	No	No	Yes	Yes	?	No	Med
de Vries G & van Hest RA, 2006	Yes	Yes	Yes	Yes	No	No	No	No	Med
McConkey S. et al, 2002	Yes	Yes	Yes	N/A	Yes	No	No	No	Med
Faccini, M. et al.2015	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	High

Study	Were the phylogenetic/ molecular methods clearly described?	Was the rationale behind the use of phylogenetics clear?	Was case finding a specific aim?	Were the methods used for case finding clearly explained?	Was it clear how phylogenetics aided case finding?	Was case finding performed?	Were undiagnosed cases found?	Directly translatable for use with HIV in the UK?	Quality*
Powell, KM. et al, 2017	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Med

* Papers meeting 1-3 criteria = Low quality, 4-6 = Medium quality, 7-8 = high quality

would not be required. Informed consent was received in two TB outbreak investigations[229, 239], though this was likely for participation in further interviews, rather than for the use of their TB genotype data. Patient consent to release identifiable data to cluster members was mentioned in one further investigation of TB[217], and consent to partake in HIV and TB contact tracing was suggested owing to the fact that some patients refused[189, 237].

2.5 Discussion

A range of strategies have been used to incorporate the use of phylogenetic data for case finding in the investigation of suspected or actual outbreaks of HIV and TB, with none describing its role in community transmission of hepatitis B, C or other STIs to date. These included identifying outbreaks requiring further action, characterising transmission to develop targeted screening programmes and using linked epidemiological data to provide clues to a sources identity. Most studies reported investigation of matching genotypes and epidemiological links in TB, reflecting its widespread use in routine surveillance and outbreak investigation[214, 249].

Few patient facilitators or barriers were mentioned, though it was suggested there were difficulties disclosing names of contacts to investigators, possibly due to issues surrounding privacy and stigma though this was not fully explored. This is likely due to the nature of the included papers, many being outbreak reports and focusing on methodological issues, and reporting of patient experience was not an objective the study. The gap in knowledge relating to public perceptions of the use of this technology leaves major questions relating to its feasibility. The use of routine HIV surveillance data to inform public health interventions on an individual level has been explored in terms of re-engaging diagnosed individuals into care, which has been found to be acceptable to PLWH[250]. However, with the complexity of phylogenetic analysis, and the identification of ‘transmission’, with connotations of blame and potential prosecution, PLWH and the general public may be wary of interventions guided by phylogenetic surveillance, particularly in terms of identifying contacts. Although phylogenetic analysis alone cannot confirm a transmission event between two individuals[210] its limitations may not be apparent to members of the pubic exposed to media coverage of criminal transmission

cases where phylogenetic analysis is used to support a prosecution. Such media coverage could alter perspectives of the use of phylogenetics for routine public health purposes. Alongside issues of acceptability and understanding, questions remain regarding potential adverse outcomes may result from such strategies being brought forward for public health practice if patient and public concerns or misunderstandings are not addressed.

Although HIV and TB are both stigmatised infections, HIV is arguably more so, largely due to the modes of transmission (mainly through sexual activity in HIV, as opposed to airborne spread), the historical conviction that HIV is untreatable, and the fact that it is currently not curable. Owing to this, there are distinct ethical differences in uncovering their transmission dynamics, with the potential in HIV to uncover sexual and personal relationships between individuals, and possibly reveal undisclosed sexual behaviours[169]. These differences require careful thought when considering whether methods described within this study investigating TB may be adapted for use in HIV. The use of linking epidemiological data to identify locations of TB transmission for screening for example, could potentially be adapted to identify sites of transmission associated with anonymous sex (e.g. sex parties, saunas) for targeted testing, and mapping of HIV transmissions to venues has been performed using phylogenetic analysis, suggesting this is practically possible[251, 252]. However, as these venues are viewed as places to relax and have fun, individuals may perceive testing in this environment as inappropriate[253]. Similarly, to adapt the strategy of investigating discrepancies between epidemiological and molecular data, detailed sexual network data would be required. The collection, storage and analysis of this may appear to intrude on individual privacy, and may even lead to reluctance to disclose contact information.

Bringing such approaches forward for public health use in HIV requires consideration of potential benefits, logistical factors and impact of interventions, alongside patient and public perceptions, barriers and facilitators. The success of molecular epidemiological methods for identifying potential contacts cannot be determined from this review, as only studies where contact tracing or screening was actually attempted and therefore considered feasible were included. Indeed, previous evaluation of TB strain typing in the UK has suggested little difference in yield of contacts, or diagnostic delay in cluster members undergoing cluster investigation and cluster members that were not investigated[254]. In addition, other potential markers of success of phylogenetically led

interventions, such as re-engagement into care and prioritising treatment of clustered cases were not addressed. Evaluating the potential public health impact of these approaches for HIV requires consideration. Performing a randomised controlled trial in situations where rapidly growing clusters of infection suggest outbreaks is problematic both ethically and logistically, given the differences between populations and unknown frequency of such outbreaks in the absence of widespread real-time phylogenetic surveillance. Evaluation of public health strategies requires triangulation of quantitative, qualitative, and ideally modelling results, and predicting the impact on HIV transmission in this context provides several challenges. The intended outcomes include a reduction in HIV transmission, and though reduction in prevalence may be used to assess HIV prevention interventions[255], with the introduction of PrEP, and current focus on scaled up HIV testing[256], determining the additional impact of phylogenetically led interventions may be challenging. In addition, modelling of these interventions, while crucial in development because of the ability to reveal previously unexpected outcomes, is unlikely to fully reflect the real-life impact of these strategies. Given the questions that remain relating to public perceptions and acceptability, as a high level of patient involvement is required in terms of sequence sampling density and in participation with proposed interventions, acceptability and engagement will determine how effective these approaches may be. Similarly, the perceptions and engagement of healthcare professionals performing these interventions will factor highly in their successful implementation and requires exploration.

Due to the nature of the study question, with no criteria for specific study design, articles identified through the electronic searches were of highly variable in terms of design and quality. As a result, studies may have been excluded that did in fact utilise sequence data to guide case finding if this was not clearly described. This may not be reported due to the objectives of the paper, or as we found in four investigations, the scientific methods and contact tracing investigations may have been reported separately. Several excluded papers that suggested the use of molecular epidemiology as a facilitator of case finding were excluded, as it was not made clear exactly *how* these methods were used; through excluding these papers, potential barriers and facilitators described may have been missed. Similarly, unclear terminology, or inadequate explanation of steps that were involved in a ‘cluster investigation’ mean that papers that did in fact utilise sequence data for case finding may have been excluded. This means the papers included in this study

may suggest a higher yield of identified contacts per cluster investigation than may be the case in real-life surveillence[254]. Within this study, an adapted quality assessment tool was used to assess the included papers, in order to address the specific study questions; what are potential methods, barriers and facilitators. Quality was therefore assessed by the clarity of the studies in describing the use of sequence data for case finding, and although this achieved the stated aim of identifying the key issues, the included studies may not be reproducible or methodologically high quality.

2.5.1 Implications for practice and future research

Future developments in the field of phylogenetically led interventions in HIV require input from current practice, public health teams and policy makers and further research. Practically, for such strategies to be used, pathways relating to the collection, processing, analysis and linkage of surveillance data require streamlining to ensure timely receipt of data required for public health action. This will impact not only on physicians submitting such data, but public health teams themselves, and the feasibility of this is yet to be determined. The processing of HIV resistance tests for clinical use may be achieved fairly quickly, but development of pathways for the subsequent analysis and linkage of data require additional time, staff and resources.

The recent introduction of GDPR has changed how personally identifiable data is processed, and provides the opportunity for patients to 'opt-out' or their data being used for any purposes other than individual care. However, data processed for the diagnosis of communicable disease and management of outbreaks of, or exposure to communicable diseases are exempt from this. Currently in the UK patient-level data used for surveillance purposes is collected and processed by PHE under section 251 of the NHS Act, 2006, as documented in schedule 19 of the Data Protection Act (updated as a result of GDPR), and is confidentially handled in compliance with the Data Protection Act and Caldicott guidelines. Patient information however may be shared with other healthcare professionals to control or prevent spread, under the 2002 Control of Patient information regulations supported by Section 251, without consent. These actions, however, must be determined proportionate by a Caldicott review panel and require review by the Office of Data Release for Public Health England meaning collaboration between bodies, with

evidence of potential benefits and consideration of privacy and security issues are key to the development of plans for phylogenetically led interventions.

Future research requires consideration of current epidemiological trends, modelling of potential benefits and patient and public acceptability. Alongside dedicated research pathways and involvement of key stakeholders and community representatives, there is a need to capture patient experience and explore and report on barriers and facilitators in practice, where anonymised or deanonymised phylogenetic analysis is used to facilitate patient level investigations in HIV, including management of outbreaks or use within the legal system. Though its use in this context may be limited at present, this could provide insight into real-life understanding and concerns, and provide guidance on how phylogenetic surveillance strategies should be adapted to ensure safe and acceptable practice.

2.6. Conclusions

Several methods using sequence data to facilitate contact tracing were identified, which could potentially be used to facilitate case finding of HIV in the UK, though initial studies are required to determine how successful these could be. There are few data on barriers or facilitators in the literature, and prior to the piloting of such interventions, exploration of the ethical issues surrounding the use of sequence data for this purpose, including legal implications, data security, privacy and consent, is urgently required. Further work to determine how these technologies could acceptably be used and how to address any potential patient barriers is needed to ensure the correct balance of benefit vs. harm is maintained.

Chapter 3: Determining the sources of transmission in a UK HIV positive cohort: a longitudinal phylogenetic analysis

Chapter purpose and summary

Given the uncertainties, including in the effectiveness of phylogenetically guided public health interventions, exploring strategies within an ethically approved research setting on a local scale would provide invaluable pilot data in order to inform more widespread interventions. The Brighton HIV cohort offer an ideal pilot population, being a fairly large centre, well engaged in research in which the majority of the local HIV positive population attend, with a presumption that a large proportion of transmission occurs locally. A previous analysis performed in order to identify factors associated with transmission found the majority of recent infections had no identifiable source, suggesting these were undiagnosed or unsampled. This study uses updated methods to try and determine whether more sources may be identified phylogenetically and whether most transmission occurs locally or is more widespread, providing evidence as to whether or not this would be a suitable pilot site for phylogenetically guided interventions.

3.1 Abstract

Background

Understanding the sources of infection within a population is key to the development of appropriate preventative interventions. A previous phylogenetic analysis found 74% of recent HIV infections (RHI) within Brighton, UK, had no identifiable likely source, suggesting these may be undiagnosed. We aimed to identify sources of RHI more accurately within this HIV cohort, to measure the extent of non-local transmission, and to determine whether real time phylogenetic surveillance may be informative within this population.

Methods

Subtype B sequences were retrieved from the Brighton population (predominantly men who have sex with men (MSM)), diagnosed 1981-2015 (n=1,840) and the most similar UK and global sequences were obtained. A maximum likelihood tree was built in RAxML (GTR + Γ), with dated phylogenies reconstructed in BEAST.

Demographic and clinical data were collected for Brighton patients, including available CD4 and viral loads, STIs, AIDS diagnoses and antiretroviral history. RHI were identified using testing history and serological markers.

Likely sources to RHIs were identified according to an algorithm considering phylogenetic and clinical data at transmission. *Potential sources* were chronic HIV infections linked to RHI, but undiagnosed at the estimated time of transmission and for which the direction of transmission could not be determined were considered

Results

360 RHI were identified, for which a likely source was identified for 178 (49%); 157 had a single most likely source, 21 had two equally likely. 93.6% of sources were male, and 92.3% of transmissions were between men who have sex with men. 64.3% of sources were local, with 35.0% from elsewhere in the UK. There was one international source.

A further 129 RHI (37%) were linked to *potential sources*, most of which appeared to be undiagnosed at the time of transmission.

Conclusions

A combined phylogenetic and clinical approach identified a potential source for the majority of RHI in this population, with a significant contribution of non-local sources, suggesting phylogenetic surveillance with the inclusion of national data may be useful to identify real-time trends in transmission within the Brighton cohort.

3.2 Introduction

Understanding the sources of HIV transmission is a crucial component of prevention planning and interventions, as highlighted in the Joint United Nations Programme on HIV/AIDS (UNAIDS) 'know your epidemic, know your response' strategy[257]. On a national scale, mathematical models, most frequently the modes of transmission model (MOT), are often used to identify sources[258], however limitations exist owing to the simplicity of the model including the assumption of closed geographical borders, and the use of representative risk and behavioural data.

Phylogenetic analysis combined with epidemiological modelling offers another route to identifying sources of HIV transmission. These methods have been used to estimate the contribution of factors including stage of infection to onward transmission[88, 259, 260]. Recent statistical approaches incorporate phylogenetic data with timing of infection extrapolated from CD4 counts and data on recent infection combined with knowledge of incidence and prevalence of infection to assign a probability of transmission between any two pair of individuals[261]. However, estimating date of transmission using CD4 decline is problematic due to the wide variation of rates of decline between individuals.

An investigation to determine the sources of infection in MSM in Brighton, UK, used a combined phylogenetic, clinical and epidemiological approach to understand the impact of recent HIV infection (RHI) and other risk factors on likelihood of transmission[88]. Here, individual level data to determine infection dates were used in combination with phylogenetic linkage to identify sources of infection for 159 RHI. Data from the source at the estimated transmission date were then analysed to identify factors associated with transmission. Only 26% of RHI had a likely source identified, suggesting the majority of

new HIV infections were probably acquired from individuals who were unsampled or undiagnosed. Though only 24% of HIV infections were undiagnosed in the UK at the time of this publication[262], they account for a disproportionate number of onward transmissions, suggesting these results to be plausible. These findings suggest that within this population, a group with high rates of viral suppression and engagement in care, focussing contact tracing efforts on newly diagnosed individuals to identify these presumed undiagnosed sources may be a crucial component of prevention. The estimated contribution of undiagnosed sources on transmission has ranged from 22-82% between studies[259, 260, 263, 264, 265, 266]. In the Brighton cohort analysis, the classification of a source as being 'undiagnosed' was based on the absence of an identifiable source within the cohort. Non-local sources were not accounted for, in part due to the presumption that this is a fairly 'closed' cohort with 88% of the local diagnosed population attending this single centre for care. Non-MSM sources were also excluded, which, due to possible misreporting of men with MSM behaviour as heterosexual[9, 169] may have resulted in fewer sources being identified. In addition, the clustering methods to identify sources enforced short genetic distances and high bootstrap support values, likely excluding transmissions from chronically infected sources. This is due to the fact that resistance tests have routinely been taken at diagnosis since 2003, and if transmission occurs from a source that has been diagnosed and off treatment for a prolonged duration, as was typical at the time of the study if the patient maintained an adequate CD4 count, the transmitted virus will have evolved, increasing the genetic distance between the sampled source and recipient viruses. As a consequence, the contribution of undiagnosed infection here is unclear.

Advances in computational speed mean that time-resolved phylogenetic approaches can now be extended to larger datasets. Transmission times can be estimated even in the absence of seroconversion dates[183]. In addition, improved knowledge of factors associated with HIV transmission mean clinical data of proposed sources may be informative in directly estimating transmission likelihood; the previous analysis classifying transmission events on phylogenetic parameters and dates of infection alone. We developed a new method for identifying sources of infection to RHI using time resolved phylogenies combined with robust longitudinal clinical data. We included non-local and non-MSM sequences to ensure all possible sampled sources were accounted for.

Finally, we discuss whether phylogenetic surveillance may be useful in guiding prevention at a local level according to our findings.

3.3 Methods

3.3.1 Population

Brighton has the highest prevalence of HIV outside London, at 8.02/1,000 of the general population aged 15-59[91]. Unlike London, Brighton has one central clinic that patients attend for care, and the cohort predominantly consists of white MSM. Unlike most major centres in the UK, the vast majority (89%) of the local diagnosed population attend this single site for HIV care. There is a high coverage of resistance testing amongst new patients and clinical and demographic data for all patients is automatically uploaded, with manual backdating since 1998.

All subtype B HIV *pol* gene sequences taken from patients from Brighton until 2016 were obtained from the UK-RDB (http://www.hivrdb.org.uk/). Where multiple tests were available, the earliest sequence was used, linked to the date of sampling and clinic number. Sequences were received aligned in a fasta file with corresponding clinical data in a separate CSV file.

In parallel, data were downloaded for all patients logged as having attended the centre until the end of 2015 from the Brighton clinical database as listed in table 5. Clinical variables were stratified into three-month study intervals and demographic variables aggregated to ensure a minimum of 20 patients shared the same characteristics to maintain anonymity. Data were then tabulated to provide a list of variables for each individual for every three-month calendar period. For those diagnosed with recent HIV infection, status was considered as recent for six months following the estimated transmission date (ETD - see below) after which they progressed to chronic infection. These clinical data were linked to the corresponding sequence, where available, using local clinic numbers, which were removed prior analysis.

Identification of recent HIV infection

Recent HIV infection was defined as diagnosis within 6 months of infection based on clinical test results. For each RHI case identified, an ETD was calculated as the midpoint of the earliest and latest possible date of infection, as determined by the assay used to identify RHI, or date of last negative test, corrected for the window period of the test used. Assays used to determine recency were an evolving serology (p24/antibody), evolving western blot and positive RITA testing.

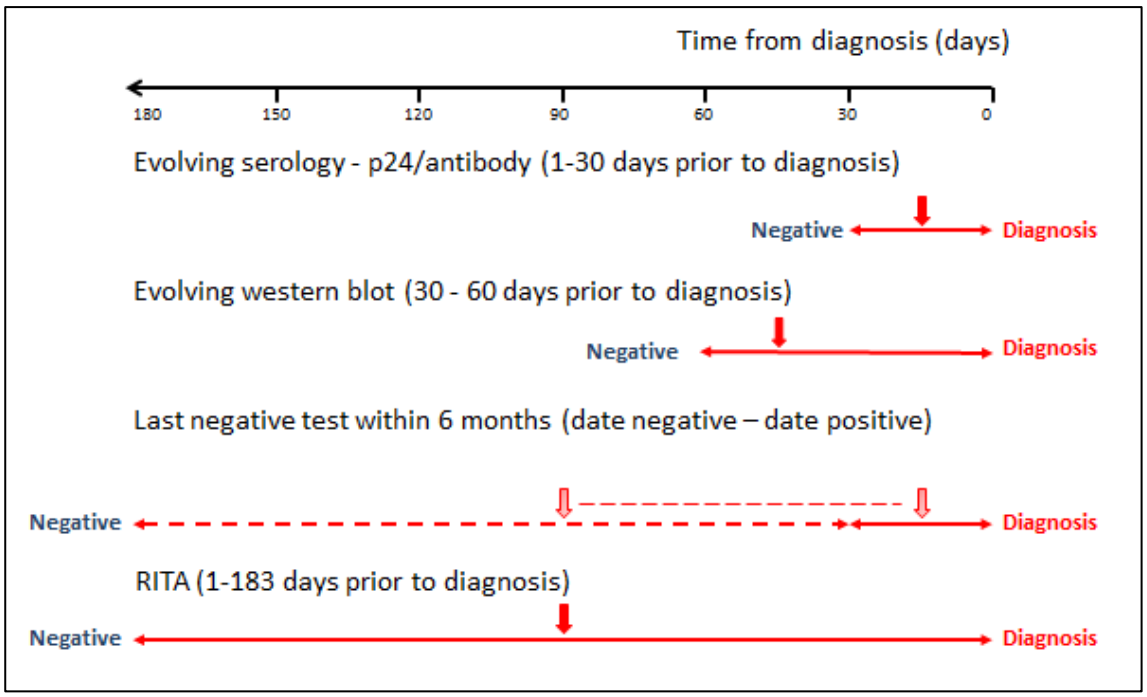


***Figure 12*: Calculation of the estimated transmission date using the midpoint of the window period for each method of identifying recent infection** ⬇ = Estimated transmission date

Ethical approval was granted on 17th March 2016 by the Proportionate Review Sub-committee of the London - Queen Square Research Ethics Committee, REC reference: 16/LO/0539.

Table 5: Clinical and epidemiological data retrieved for Brighton patients

Data collected	Variables
Sex	Male, female
Sexuality	Heterosexual, homosexual, bisexual, other/not stated
Year of birth	
Ethnicity	White, Black*, Asian, other
Migration data	Country of birth, date of arrival in UK, country of infection (by continent if not UK)
Follow up data	Dates of attendance, date of transfer from/to other sites, dates 'lost to follow up'
Postal region	Brighton, Southeast, London, other
Clinical data	Date of diagnosis, date of death, ART start/stop dates. All CD4, viral loads (VL), sexually transmitted infections (STI) and AIDS diagnoses
Seroconversion data	RITA results, evolving serology, western blot, date of negative test

*Combined black African, black Caribbean ethnicities owing to small numbers

3.3.2 Background sequences

The first sequence for every individual with subtype B HIV-1 within the UK-DRB was retrieved and linked to clinical and demographic data held by the national HIV/AIDS reporting system (HARS), excluding individuals from Brighton. In addition all subtype B partial pol sequences (HXB2 positions 2253-3700) or fragments of a minimum 700 base pairs were retrieved from the Los Alamos HIV Sequence Database (LANL), linked to country, sample date and clinical information where available (http://www.hiv.lanl.gov/). The Basic Local Alignment Search Tool (BLAST) in Geneious V 1.8.2 was used to identify the 10 closest UK and LANL sequences to each Brighton sequence, and duplicates were removed. Using a custom R script, primary drug resistance mutation sites taken from the 2013 International AIDS Society list[267] were stripped from all sequences used in the analysis. This was to reduce the chance of ascribing similarity to sequences from individuals who have obtained similar drug resistance patterns due to previous ART regimens.

3.3.3 Phylogenetic methods

Maximum likelihood (ML) trees were constructed in RAxML[268] under the generalised time reversible (GTR) model of nucleotide substitution and gamma (Γ) distribution of rates. ClusterPicker was then used to identify all clusters within trees with a genetic distance cut-off of 4.5%, and ClusterMatcher to identify all of these clusters containing at least 1 Brighton sequence[154]. Clusters containing Brighton sequences were extracted and the sequences contained within those clusters made up the final sequence dataset. Clusters were pooled for analysis, with all sequences from the same cluster contained within the same dataset and each dataset comprising no more than 300 sequences, owing to limitations in terms of sample sizes of the phylogenetic programmes used. Demographic models were initially compared in a subset of datasets using path sampling and stepping stone sampling. The logistic, constant, exponential and Skyride demographic models were evaluated using BEAST v1.8.3.

An uncorrelated relaxed molecular clock[148] with the SRD06[269] model of nucleotide substitution and Skyride[270] demographic model was selected to run all sequence datasets. Analyses were run in duplicate for 100,000,000 generations, sampling parameters every 10,000 generations. Resultant files were combined using the programme Logcombiner, a 10% burn-in was removed and resampled to leave 1000 states. Effective sampling size (ESS) values were inspected (the number of independent draws that the Markov chain is equivalent to, providing confidence in the posterior distribution of specific parameters) and maximum clade credibility (MCC) trees were constructed using TreeAnnotator v1.8.3 (http://beast.bio.ed.ac.uk/TreeAnnotator).

3.3.4 Identifying sources of infection to RHI

A custom R script was created to extract sister tips (sequences that share a common node) for each RHI, alongside branch lengths from which the mean node age could be calculated, by inputting the MCC tree and sequences. Genetic distances were calculated under a K80 substitution model for pairs of sequences within clusters where more than one sister tip was associated with an RHI, and the tip with the shortest genetic distance was selected. These were combined with clinical data for the candidate source, including dates of positive and negative tests, dates of death, and treatment status, CD4, viral load

and AIDS and STI diagnoses at the ETD using Microsoft Access. Clinical data were analysed to determine whether a transmission to the recent infection was possible (i.e. the proposed source was alive, diagnosed or likely to have undiagnosed infection determined by dates of negative testing and CD4 at diagnosis, with a high viral load at the ETD). Potential transmission pairs were then reviewed in the tree. The date interval of likely viral divergence within the candidate *source* was estimated using the 95% highest posterior density (HPD) interval at each node and posterior probabilities were reviewed to determine the likelihood of a transmission event. If these data did not support transmission, or support for the node was poor, other sequences within a well-supported cluster were investigated with the same process. In addition, sources were rejected if branch lengths or HPD intervals exceeded 5 years given the increasing possibility of transmission from an unsampled source, unless there was very strong evidence to support a transmission event. Genetic distances were used only as an initial step in investigation of clusters with multiple tips, and were not used to assign transmission events, in order to reduce bias towards recently infected individuals[145].

RHI were categorised as having likely, potential, or no identified transmission source. Sources were considered 'likely' if clinical data was supportive, 95% HPD intervals were relatively narrow and overlapped with the likely time of HIV acquisition in the candidate source and the branching pattern and node supports corresponded. 'Likely' sources could be undiagnosed if they were likely infected before the EDT of the RHI, as determined by clinical data at diagnosis, including CD4 and RITA test results, and the 95% HPD interval corresponded with their likely HIV acquisition window, and preceded that of the RHI.

RHI were considered to have 'potential sources' if transmission may have occurred, but uncertainty was introduced through clinical or phylogenetic parameters. These included cases where direction of transmission could not be determined with confidence, RHI were in large clusters with poor branch support values resulting in multiple potential sources, or if intermediaries could not be excluded owing to broad HPD intervals or long branch lengths, often encountered where there was a long delay between diagnosis and sequencing.

If clinical or phylogenetic data excluded a transmission event, recent infections were classified as having no potential transmitter, suggesting the source may be unsampled.

Source clinical data supportive of transmission event;

- Alive and diagnosed, or CD4 trajectory +/- timing of AIDS diagnoses suggests source had undiagnosed HIV at the ETD
- High viral load at ETD
- ART naïve or treatment interruption at ETD
- STI diagnoses

Non-supportive;

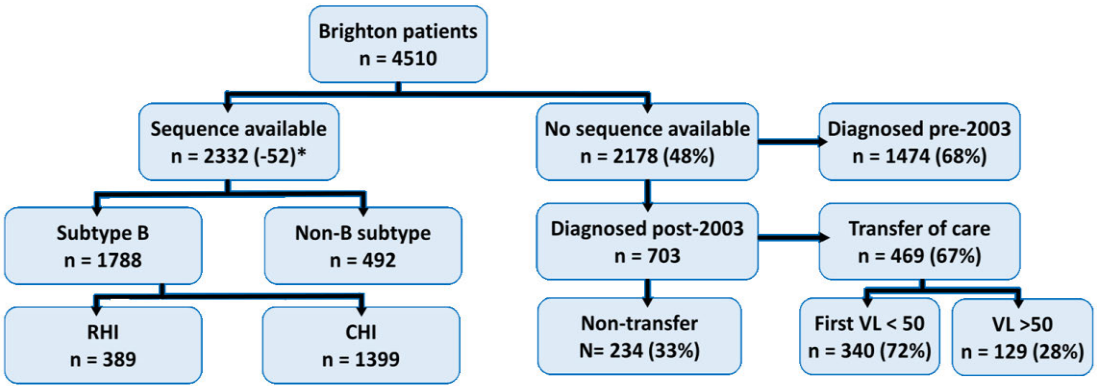
- Low viral load at ETD
- Female source to female recipient if sexual transmission
- Source not resident in UK at ETD
- Source unlikely to have acquired HIV prior to the ETD

3.4 Results

3.4.1 Study sample

Of 4,510 patients logged as having attended for HIV care at the RSCH HIV centre pre-2016, 2,332 had a sequence available, of which 1,840 were subtype B (79% of those with a sequence). Of these, 52 were of inadequate length for analysis leaving 1,788 sequences. With the addition of 6,106 UK sequences, and 2,495 international sequences, 10,389 sequences were included in the full dataset.

Of patients without an available sequence, 1,474 (67.7%) were diagnosed before baseline genotyping was introduced into routine practice in 2003. Of the 704 patients diagnosed post-2003, 340 (48.3%) had their first viral load recorded as undetectable, suggesting they had transferred in to Brighton when virologically suppressed and that their contribution to transmission would be minimal after entering care in Brighton and thus being classed as a ‘Brighton patient’, though transmission may have occurred previously when attending another site, when classed as a ‘non-Brighton patient’. Of the 363 with a detectable first viral load, 129 (35.5%) had documentation of transfer in from another site. Overall, of those diagnosed in Brighton post 2003, 89% had a sequence available.



*sequences too short to analyse so excluded

***Figure 13*: Sequences available from patients that had attended for care at Brighton**

3.4.2 Recent HIV infections

389 of the 1,788 subtype B Brighton individuals were identified as having a recent infection at the time of diagnosis, diagnosed between 1985 and 2015 and sequenced between 1997 and 2015. 29 appeared to have transferred their care into Brighton after diagnosis, leaving 360 RHI diagnosed in Brighton. 61 (16.9%) were identified through evolving serology (p24 antigen/antibody), 3 (0.8%) through evolving western blot, 117 (32.5%) through dates of last negative test, and the remaining 179 (49.7%) were identified through positive RITA testing. 357 (99.2%) were male, and 353 (98.1%) acquired their infection through sex with men. 15 (4.2%) self-reported likely transmission abroad.

3.4.3 Sources of infection

3.4.3.1 'Likely' sources

Of 360 RHI, a likely source was identified for 178 (49.4%); 157 having a single most likely source and 21 having two equally likely sources. Of the 157 single sources, 147 (93.6%) were male and 3 (1.9%) were female. 141 (89.8%) single sources were white. 145 (92.3%) transmissions were between MSM and 6 (3.8%) were between men and women. 50 (31.8%) of these single likely sources were undiagnosed at the estimated time of transmission.

Table 6: Locations of single likely or potential sources to recent infections from Brighton

	Single likely source (%)	Potential source (%)
Local: Brighton	101 (64.3%)	48 (37.2%)
UK	55 (35.0%)	86 (59.7%)
global	1 (0.6%)	5 (3.5%)
total	157 (100%)	129 (100%)

3.4.3.1.1 Local likely source

Of the 157 single likely sources of Brighton RHI, 101 (64.3%) were also from Brighton, meaning overall, 28.1% of RHI in Brighton had a likely local transmission source. All local likely sources were male, 96 (95.0%) white, with a mean age of 32 (range 20-62 years). All transmissions were between MSM apart from one transmission from a heterosexual male to a female RHI.

37 (36.6%) of the 101 local likely sources were undiagnosed at the ETD and 8 sources (7.9%) were themselves in RHI at the time of transmission. 11 (10.9%) had a concomitant STI at transmission. Of the 64 diagnosed local sources 6 were undergoing a treatment interruption, 4 were reported as being on treatment but viraemic, and the remaining 54 were ART naive.

Table 7: Characteristics of 'likely' sources to recent infections from Brighton

	Brighton source, n=101, (% total)	Non-Brighton UK source, n=55 (% total)
Sex		
Male	101 (100%)	47 (85.5%)
Female	0 (0%)	3 (5.5%)
Unknown	0 (0%)	5 (9.1%)
Ethnicity		
White	96 (95.0%)	46 (83.6%)
Black	1 (0.9%)	2 (3.6%)
Other	4 (3.9%)	0 (0%)
Unknown	0 (0%)	7 (12.7%)
Mean age in years (range)	33 (20-62)	36 (20-60)
HIV status at EDT		
Undiagnosed	37 (36.6%)	13 (23.6%)
Undiagnosed in RHI	5 (5.0%)	
Diagnosed	64 (63.4%)	42 (76.4%)
In RHI	3 (3.0%)	
Chronic, untreated	51 (50.5%)	
Chronic, on ART	4 (4.0%)	
Chronic, ART interruption	6 (5.9%)	
STI diagnosis	11 (10.9%)	
Lost to follow up at EDT	2 (1.9%)	

3.4.3.1.2 Non-local likely source

55 (35.0%) of the 157 single likely sources were from elsewhere in the UK meaning overall, 15.3% of recent HIV infections in Brighton had a non-Brighton UK source.
Of those with a reported gender (50/55), 47 (94%) were male and 3 (6%) were female. Of those with a reported ethnicity (48/55) 46 (95.8%) were white and 2 (4.2%) were

black. There were 45 (81.8%) male to male transmissions, and 5 (9.1%) heterosexual transmissions (3 female to male, 2 male to female). 13 (23.6%) of the likely sources were undiagnosed at the estimated transmission point.

The largest single geographic group of non-Brighton sources attended for care in London (40.0%, n=22). The remaining 60% were spread throughout the UK. Due to incomplete data from this group, stage of infection was difficult to determine for most sources, and STI data were unavailable.

In addition, one individual with recent HIV was found to have a likely source from outside of the UK. This individual self-reported acquiring their infection in that world region.

3.4.3.1.3 Two equally likely sources

21 further RHI had two equally likely transmission sources. In 15 cases (71.4%), both sources were from the same region (12 both Brighton, 3 both from the same UK region) suggesting they may represent part of the same transmission chain.

3.4.3.2 Potential sources

Of those without a likely source, potential sources could be identified for a further 129 (36.1%) individuals, of which 48 (37.2%) sources were from Brighton, 76 (58.9%) from elsewhere in the UK, and 5 (3.9%) from outside of the UK. Most of potential sources were likely undiagnosed at the ETD and direction of transmission could not be determined, for example due the 95% HPD interval overlapping with the period of acquisition of the RHI. In two cases, the phylogenetic data were strongly suggestive of transmission but the potential source was diagnosed and virally supressed at the ETD suggesting an unsampled third party or data error.

Table 8: Locations of potential sources to recent infections from Brighton

	Potential source (%)
Local: Brighton	48 (37.2%)
UK	86 (59.7%)
global	5 (3.5%)
total	129 (100%)

3.4.3.3 No source identified

53 individuals had no possible source identified. Of these, all were male, and 96.2% were MSM. In this group, a higher proportion were diagnosed pre-2003 than those with a likely/potential source (43.4% vs. 15.3%, $p<0.001$) and a higher proportion were of black ethnicity (7.5% vs. 0.7%, $p<0.001$), though absolute numbers with black ethnicity were small, owing to the characteristics of the Brighton population.

Table 9: Characteristics of RHI with an identified source vs. RHI with no identified source

	RHI with a source (%) (n=307)	RHI with no source (%) (n=53)
Male	304 (99%)	53 (100%)
MSM	299 (97.4%)	51 (96.2%)
Female	3 (1%)	0 (0%)
Mean age	33.8	34.4
Age range	16-68	19-59
White ethnicity	294 (95.8%)	48 (90.6%)
Black ethnicity	2 (0.7%)	4 (7.5%)
Other ethnicity	7 (2.3%)	1 (1.9%)
Unknown ethnicity	4 (1.3%)	0 (0%)
Average year of diagnosis	2007	2003
Diagnosed pre-2003	47 (15.3%)	23 (43.4%)
Diagnosed post-2003	260 (84.7%)	30 (56.6%)
Infected abroad	10 (3.3%)	5 (9.4%)

3.5 Discussion

With the inclusion of UK and global sequences and an extended duration of follow up, we identified a likely or possible source of infection for the majority of our population using an individual level approach combining detailed clinical data and phylogenetic parameters. We were able to identify a source fairly confidently for 49% of 360 RHI owing to clear direction of transmission, clinical data suggesting a high transmission risk and supportive phylogenetic parameters. For a further 36%, we identified one, or several potential sources, where clinical or phylogenetic parameters could not confidently determine direction of transmission, nor exclude an unsampled intermediary.

A significant proportion of sources were found to be non-local, with the largest proportion of non-local sources (40%; 14% of all single likely sources) originating in London, the closest major city. Both Brighton and London have prominent 'gay-scenes' and transmission from London residents visiting Brighton to socialise, or vice versa are both very plausible scenarios. Other non-local sources were geographically spread throughout the country, and it is feasible that these may have also resulted from social visits to these prominent cities. Although centre of care is used in this analysis, care may not equate to location of residence, however given the high proportion of Brighton based patients attending the local centre (89%), the proportion of misclassified non-local sources is likely to be fairly small. These data therefore may not be fully representative of the extent of local vs. non-local transmission. However, only one of the 39 Brighton RHIs living in London was linked to a London source suggesting mixing between Brighton and non-Brighton residents. Within the Brighton cohort, where STI data were available, we found high rates STIs amongst sources, both reflecting the heightened HIV risk of concomitant STIs[23] and high risk behaviour within this cohort.

Given the inclusion of all available similar UK and global sequences, and the extended duration of follow up, with an estimated 89% coverage of sequencing post 2003 we would expect to have the majority of transmission sources included within this analysis given 86% of RHI were diagnosed after 2003. This suggests our results may provide a reasonable representation of transmission within this cohort despite the large proportion of patients who were recorded as attending for care without available sequences.

However, this approach identified a higher proportion of likely diagnosed sources than we would expect. Although previous phylogenetic studies have shown up to 77% of sources to RHI were diagnosed at transmission[263], and similar transmission rates have occurred between serodiscordant couples on or not on ART within one population[271], given this population has both high engagement to care rate and treatment coverage, we would anticipate a smaller proportion. Our findings may be in part due to the timescale of the follow up; mapping transmission from 1995 until 2015, prior to universal ART when many individuals diagnosed with HIV would have a significant delay from diagnosis to treatment, and treatment interruption was accepted in practice. However, estimates for transmission over the final few years of analysis may overestimate the contribution of diagnosed individuals due to the exclusion of sources who may have been diagnosed and sequenced subsequent to retrieval of the dataset. Nevertheless, although 109 (69%) of 'likely sources' were diagnosed at the estimated transmission date, overall only 30% of RHI had a diagnosed likely source. Using in depth clinical data at the ETD allowed us to assign a likelihood of transmission according to individual situations and was highly informative when the source was diagnosed. However confidently identifying undiagnosed sources was made difficult due to a lack of clinical data at the transmission date, particularly in situations where there was a long delay from transmission to diagnosis and where consequently undiagnosed sources were classified as 'potential' rather than 'likely'. Therefore, 'potential sources' may well reflect the majority of undiagnosed transmission within this population, which in an era 90-90-90 goals, and the introduction of universal ART at diagnosis are an increasingly important target for public health interventions.

The contribution of recent HIV to transmission appears to be lower in this analysis than previously identified within this cohort (8% vs. 26% amongst local sources). Recent HIV infection increases transmissibility of HIV[89] though due to its relatively short duration, it's impact on transmission may be reduced[272] and estimates of the contribution of recent infection on population level transmission have varied widely[90, 183, 273, 274, 275]. Our more conservative estimates may be in part due to the disregard of genetic distances within our source identification algorithm and acknowledgement of progression of RHI to chronic infection[276], with an increased proportion of time for most individuals in the chronic untreated stage of infection due to the long follow up period. However recent phylogenetic studies in similar populations in the USA and the Netherlands have

suggested that almost half of transmissions occur within a year of the source acquiring HIV[259, 260] and it is likely our methods are an underestimate due to the proportion of undiagnosed 'potential' sources who may have transmitted while themselves having RHI.

Phylogenetic analysis of *pol* cannot confirm transmission between two individuals and a phylogenetic tree does not equate to a transmission tree, so although the data presented is suggestive, firm conclusions cannot be made in regards to individual transmission events, as exemplified by the 21 RHI for whom two sources were deemed equally likely. Unsampled intermediaries cannot be excluded within our transmission pairings; almost half of Brighton attendees pre-2016 had missing sequences, suggesting this is a likely possibility. However, this is likely to be an overestimation of the number of local sequences missing; attendances logged on the local database may relate to non-local individuals that have visited and collected prescriptions, or been admitted acutely but continued follow up elsewhere. Further analysis of these missing sequences revealed that the majority were diagnosed earlier than 2003 and the introduction of baseline genotyping, or had transferred care and were on treatment. As the majority (86%) of RHI were diagnosed post 2003, after which sampling density was approximately 89%, we expect the effect of these missing sequences on our results to be significantly reduced. However, even within pairs that had data supporting transmission, without additional contact tracing data, the relationship between viruses cannot be confirmed. The potential for misclassifying transmission pairs was highlighted by two cases which by our method appeared 'likely' but the sources had an undetectable viral load at the estimated transmission date. Had viral load data been missing that quarter, potentially these may have still been classified as likely. In addition, some available clinical data were conflicting, for example, transmission between MSM and heterosexual women, and the identification of UK sources to individuals that were reported as being infected abroad. In the UK, between 9.3 and 28% of MSM report one or more female sexual partners in the past year, with 44.3% of gay identifying MSM reporting attraction to the opposite sex[277]. In parallel, analyses estimating country of infection by date of UK arrival and CD4 decline have suggested clinic-reported country of infection may overestimate the contribution of infections acquired abroad, due to the underestimation of risk in the UK, underreporting of risk behaviours, and complex travel and migration histories[278]. Therefore, as there was uncertainty in the relevance or reporting of these factors, these

data were considered to refute transmission only in cases where other data was less supportive.

The identification of such a high proportion of likely or potential sources within this cohort using epidemiological data and baseline HIV sequences suggests that real-time phylogenetic surveillance of this group may be informative of future trends in transmission if analysed in conjunction with national data. However, the exact role of the 'potential sources' remains unclear. Phylogenetic methods have been developed that are able to assign a probability of direct transmission between individual viruses and potentially direction in the presence of appropriate clinical and epidemiological data, which may clarify the role of the 'potential transmission sources'. However, identifying sources by examining transmission events with individual longitudinal data allows for assessment of transmission events using parameters that could not yet be incorporated into such models; individual CD4 responses over time, viral load levels, presence and frequency of STIs suggesting high-risk behaviour, and non-adherence to therapy. Combining this more 'clinical' approach to source attribution, alongside statistical approaches previously described could further clarify the roles of these 'potential' sources. Characterising undiagnosed sources more accurately may help guide public health interventions and possibly determine whether we may be able to prospectively predict their presence within a real-time phylogenetic surveillance network, leading to earlier diagnosis and less onward spread.

3.6 Conclusions

We identified potential sources of most RHI within a local UK HIV positive cohort, using a combined phylogenetic, clinical and epidemiological approach, showing transmission is not confined within a local population and prevention strategies require national collaboration. Further analysis is required to confirm the extent of transmission from undiagnosed sources, which in an era of universal treatment at diagnosis and strong evidence that suppressed virus cannot be sexually transmitted will be a priority for HIV prevention in the UK.

Chapter 4. Surveillance, Ethics and Phylogenetics, a literature review and synthesis

Chapter purpose and summary

The use of public health data for surveillance and action is intertwined with ethical dilemmas and considerations. With the evolution of surveillance, several guidelines on the ethical use of surveillance data have been developed, both in the UK and globally, based on experience and expert review. When considering the development of phylogenetically guided interventions, an area in which even the most effective practical methods are still unknown and possibly unexplored, understanding these guidelines and anticipating possible ethical challenges that may emerge with their use is crucial. As the use of phylogenetics for public health intervention is a very new field, our understanding of the ethical issues specific to this area is limited.

This chapter provides a systemic review of current surveillance methods and ethical guidelines for these strategies both in general and specifically in HIV, and explores literature highlighting ethical issues directly associated with the use of HIV phylogenetics, including for prosecution of transmission. The aim of this chapter is to introduce these concepts, to provide insight on what is ethically possible with phylogenetics for public health interventions. In addition, the findings of this review complement chapter 2 in the development of theoretical phylogenetically guided interventions described in scenarios developed for the subsequent study, exploring stakeholder views on the use of phylogenetics. The information explored in this chapter ensures proposed strategies are both plausible, and as ethically sound as possible at this early stage of exploration, to ensure the results are applicable to realistic future uses.

4.1 Methods

The aims of this chapter are to define key concepts in respect to surveillance and current practice, to outline current guidelines and literature concerning the ethical conduct of surveillance and public health action, and to use these concepts to synthesise a framework of ethical considerations relating to the use of HIV phylogenetics to guide intervention. The approach to conducting this review therefore comprised searches of various data sources in order to address each aspect.

To identify definitions, protocols and guidance searches were conducted within organisational websites, including the World Health Organisation, Public Health England and the General Medical Council. To identify research and published reflections on relevant ethical issues, the following searches were conducted in PubMed; (ethic* AND surveillance AND HIV); (ethic* AND phylogen* AND HIV); and (criminal* AND phylogen* AND HIV). A separate search was used to identify articles discussing criminalisation, an ethical issue identified earlier in this thesis, as some relevant articles may have been missed using terms only relating to ethics.

Titles were screened to identify relevant papers, with reference searches as appropriate. Papers were re-read, and findings synthesised according to a predetermined structure; definitions and current practice; ethical issues previously identified within HIV surveillance and phylogenetics; and how these issues may impact on the use of prospective phylogenetics to guide intervention.

4.2 Public health surveillance versus public health research

The World Health Organisation (WHO) define surveillance as "the continuous, systematic collection, analysis and interpretation of health-related data needed for the planning, implementation, and evaluation of public health practice"[279]. It is used throughout the world to guide public health practices relating not only to outbreaks and

epidemics of communicable diseases, but to the management and control of non-communicable disease, for example obesity and ischaemic heart disease. Its primary intent is to enable direct or indirect public health action, through identifying public health issues, identifying risk factors for poor health and through the design and evaluation of interventions for control.

Public health research on the other hand refers to "the attempt to derive generalisable or transferable new knowledge to answer or refine relevant questions with scientifically sound methods", as defined by the Health Research Authority. Public health practice may however also provide generalisable results depending on the conditions or populations under investigation, and methods may overlap, blurring the distinction between the two. Differentiation of the two stands most clearly in the purpose of the activity – the main outcome being to generate generalisable knowledge in research, as opposed to preventing or controlling illness in the participating community through surveillance, monitoring or development of public health services.

Public Health England (PHE), an executive agency sponsored by the Department of Health, are responsible for public health surveillance in the UK, alongside promotion of healthier lifestyles, protection of the public from potential health hazards including through health screening and vaccination, responding to public health emergencies and identifying and preparing for future public health challenges. (https://www.gov.uk/government/organisations/public-health-england/about).

PHE comprise a robust team of scientists, public health officials and researchers, that work closely with the NHS, local government and other officials to provide a national infrastructure for health protection and provide multiple services including integrated surveillance, diagnostic and reference tools for infectious disease including the development and use of genomic technology for health, investigation and management of outbreaks and promoting and undertaking research to provide evidence to support the development of interventions and policies to promote the health.

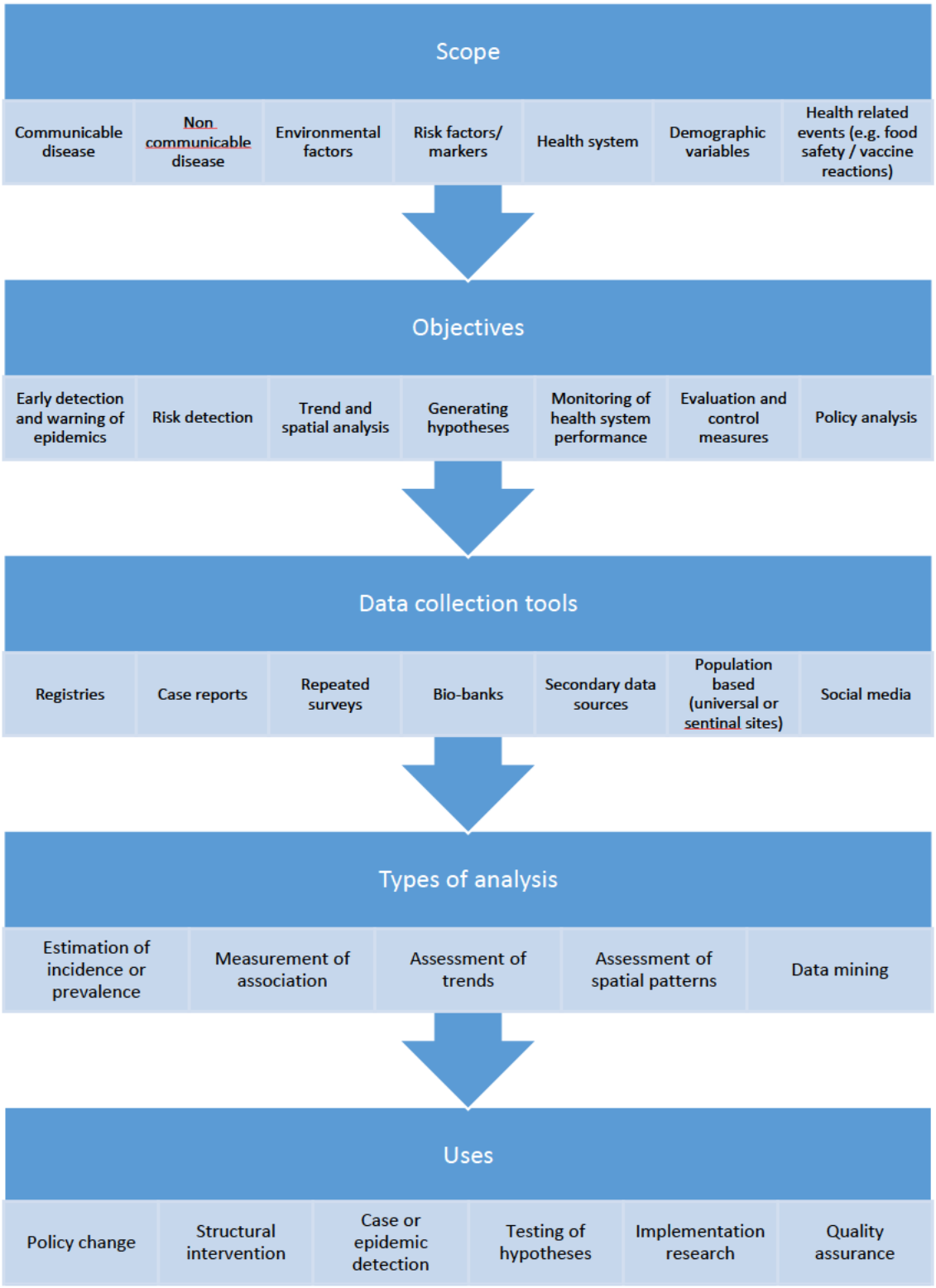


***Figure 14*: Dimensions of Public Health Surveillance. Adapted from WHO guidelines on ethical issues in public health surveillance (2007)[280]**

4.3 Surveillance in HIV

After the Centres for Disease Control and Prevention (CDC) in the USA first reported cases of PCP pneumonia in gay men in 1981[281], on-going surveillance of HIV has been carried out, initially relying on reports of AIDS related illness and a few sentinel surveillance projects to help determine prevalence of HIV, processes termed 'primary surveillance'. The introduction of second-generation surveillance in 2000 built on these systems to allow more accurate monitoring of epidemics and evaluation of responses, tailored to each country and population. This is achieved through accurate identification and monitoring not only of identified cases with HIV and AIDS, but also identification of groups at risk and behaviours associated with transmission, with an aim of concentrating resources to where they will have most impact on the prevention and management of HIV[282]. Methods utilised to reach these goals include sentinel surveillance, case reporting, collection of routine clinical, epidemiological and behavioural data and biological and behavioural probability surveys.

4.3.1 HIV surveillance in the UK

PHE hold several HIV surveillance systems addressing various aspects of HIV epidemiology and management. These include The HIV and AIDS Reporting System (HARS), Recent Infection Testing Algorithm (RITA) monitoring and the CD4 Surveillance Scheme.

The HIV and AIDS Reporting System (HARS), preceded by the HIV & AIDS New Diagnoses Database (HANDD) and the Survey of Prevalent HIV Infections Diagnosed (SOPHID), is a disaggregate dataset submitted on a quarterly basis by all UK HIV outpatient clinics. This comprises demographic information, attendance information (region of clinic, appointment type), diagnosis information (dates, risk exposure, country of infection), antiretroviral treatment histories and clinical information including CD4 counts, viral loads, AIDS diagnoses and dates of death. Its purpose is to identify risk groups, monitor clinical outcomes and quality of care, monitor effectiveness of national policies and inform public health response, with a secondary aim of informing commissioning of HIV services in the UK.

As HIV is not a notifiable disease in the UK, reporting is not mandatory and is done on a voluntary basis by clinic staff. Patient consent is not required, though they do have the option to opt-out of the reporting of their personal data. Though the data is pseudoanonymised form, due to the risk of deductive disclosure with the addition of multiple variables, approval under section 251 of the NHS Act 2006 was sought and granted for the collection of this data without explicit consent. Strict safeguards are in place within the organisation to ensure this sensitive data is protected, with processing of data performed under the Data Protection Act and Caldicott principles. Sharing of data is under the guidance of the PHE HIV and STI data sharing policy, protecting against deductive disclosure on sharing of HIV related information through measures including rounding up in denominations with small cell sizes, and prohibitation of sharing of individual level data.

4.3.2 Genetic surveillance

HIV sequences derived from routinely collected drug resistance tests are transferred the UK HIV Drug resistance Database (UK RDB) (http://www.hivrdb.org.uk/), a central repository for all tests taken in the UK, since 2001. Although the UK RDB is a research study, its primary objective at set up was to research and monitor trends in HIV drug resistance alongside research into its epidemic dynamics, therefore contributing to the surveillance of transmitted drug resistance in the UK, determined from baseline resistance. Data are collected yearly from 15 participating laboratories, and where possible linked to clinical data from PHE, UK CHIC (the UK Collaborative HIV Cohort study), the UK Seroconverter Register and CHIPS (The Collaborative HIV Paediatric Study). These data include dates of diagnosis and negative tests, CD4 counts and HIV viral loads, antiretroviral history, and epidemiological variables, coded and aggregated to reduce the risk of deductive disclosure. Data are then cleaned and anonymised, and made available to collaborators on submission of approved research applications.

The database reports on yearly rates of transmitted drug resistance, providing information on whether current first line therapies are appropriate in the context of expected resistance rates[283]. Data are also used for phylogenetic analyses, providing insight into epidemiological dynamics, however due to time considerations in the collation, preparation and cleaning of datasets, these are generally only available from a minimum

of two years prior to analysis. This means that although we may gain useful insight into recent transmission patterns using the UK RDB, the use of phylogenetics for contemporaneous surveillance of HIV transmission in the UK is not yet possible in its current setup.

4.3.3 The use of individual level surveillance data for HIV interventions

4.3.3.1 Management of outbreaks

Investigation of outbreaks are a key aspect of public health surveillance, the International Health Regulations (2005) defining surveillance pertaining to outbreaks as "the systematic, ongoing collection, collation and analysis of data for public health purposes and the timely dissemination of public health information for assessment and public health response as necessary"[284]. Outbreaks are considered to be two or more epidemiologically linked individuals suffering from the same illness; an increased incidence of infection compared to expected rates; a potential or actual event arising from contaminated food or water; or a single case of illness in specific rare diseases[285], with the former two definitions being relevant to HIV infection. Unlike much routine surveillance, due to the actions that may be required to protect the public, patient level data may be utilised without explicit consent, under Section 251 of the NHS Act 2006.

Although previously documented outbreaks of HIV in the UK have been largely identified as a result of local surveillance or investigation though contact tracing activities[17, 177, 178], identification of apparent discreet outbreaks requiring further investigation is possible at a national level, and additional assistance with outbreak investigation activities may be required from governing bodies necessitating the use of patient level surveillance data. Additional activities may include supplementary testing of pre-collected lab specimens, in depth case interviews, comprehensive partner notification, and outreach activities targeting specific venues/locations.

4.3.3.2 Surveillance data used to improve individual level outcomes

Interventions to protect health are major outcomes of surveillance; however the use of surveillance data to intervene at an individual level is non-standard owing to issues of

confidentiality, apart from in potential outbreak situations where the need to act to protect public health may be deemed large enough infringe on individual liberty.

In the days when treatment for HIV was not available, debate was held as to whether the public health benefit of surveillance activities outweighs the risk to the individual. Activists strongly opposed to use of personal data that may lead to any intervention which may pose a risk of increasing the stigma that was prevalent against affected groups, which was felt may potentially lead to a breach of human rights, with some believing this data could be used as a basis for quarantine, to contain infection spread. The use of surveillance data to guide individual interventions therefore has been limited, however with the introduction of mandatory reporting of CD4 counts and viral loads in some states in the USA since 2006, its use has increased.

Several states in America have now implemented re-engagement services based on laboratory data collected for surveillance purposes, used as a proxy for engagement into care (for example, the absence of CD4 or viral load data, or persistently high viral loads suggesting individuals may not be accessing medical care or treatment). Early examples of this use, termed 'Data to Care', include in New York City, where from 2007 the close working of individuals within surveillance, outreach and medical departments enabled identification, verification and attempted strategies for re-engagement and partner services to those who had appeared to have disengaged. Evaluation of the service found that 59% of those with surveillance data suggesting disengagement that could be located had indeed disengaged, 77% of whom expressed willingness to re-engage, with 57% of whom were successfully re-engaged into care[286]. Similarly, after expansion of the 2006 pilot, the Not in Care Evaluation (NOTICE) in King County, Washington[287], public health workers contacted patients confirmed to have disengaged, after approval from their direct health care team, identified by either an absence of lab results in the past year, or a most recent CD4 of <500 cells/cm^3 and VL of >500 copies/ml. Although assessment showed the majority of individuals classified by the system as disengaged had in fact transferred out of the area, modest, but statistically significant improvements in rates of re-engagement in those correctly identified were acheved[288].

Developing on from this, in 2009, the state of Louisiana implemented the Louisiana Public Health Information Exchange (LaPHIE), developed by the Office of Public

Health[289]. This secure bidirectional electronic health information exchange links surveillance data to electronic health records, triggering an alert to the health professional in any department if consulting with an individual thought to have disengaged with HIV care (no CD4 or VL for >12 months), or if an infant exposed to HIV engages with any type of medical care at the county's health facilities, prompting discussing and facilitation back into care. An analysis over 2 years showed 82% of those identified by the system attended for healthcare during the study period.

'Data to care' has not progressed without difficulties however; the first evaluation of NOTICE found that almost 80% of individuals identified were misclassified as disengaged when they were still in fact in care, had transferred or died leading to evaluation and improvement in reporting of surveillance data locally[290]. Another similar American pilot project undertaken in several states performed poorly due to inaccuracies and inconsistencies in reporting of surveillance data[291]. In addition, an investigation reviewing contact information within surveillance registries within the US found that incorrect telephone numbers were held for 78% of individuals, incorrect addresses for 61% and incorrect HIV care facility for 22%[292]. Although surveillance staff were able to search other registries for these data, it shows how inaccurate reporting may make intervention from these data, even if informative themselves, impossible.

Although there have been reported concerns from public health staff[293], assessment of acceptability has shown that PLWH generally support the notion of 'Data to Care'[250, 289, 294], though some voiced concerns regarding being contacted by a public health professional over their routine doctor, with provider- mediated outreach[250, 294].

4.4 Exploration of ethical issues of surveillance

4.4.1 Ethical guidelines for epidemiological studies

Since the creation of the Nuremberg code in 1947, followed by The World Medicine Association's Declaration of Helsinki in 1964, informed, voluntary consent has been the cornerstone of ethical research, alongside the prioritisation of the rights and confidentiality of the individual over those of society. The rigidity of these principles however raised concerns regarding the future of epidemiological research, particularly when relating to large study populations and the review of patient records, when informed consent may not be feasible, or may even be impossible[295, 296]. In 1981, the US Department of Health and Human Services waived the need for informed consent for epidemiological research if this was not possible, data were not recorded in a way that may be identifiable and risks to subjects were minimal, however European guidelines at this time still prioritised participant consent. In 1991, the Council for International Organizations of Medical Sciences (CIOMS) published the 1991 International guidelines for ethical review of epidemiological studies[297], which acknowledged the difference between researching groups or populations, and researching individuals, and provided guidelines on consent based around the three ethical principles identified by the 1974 Belmont report[2]; respect for persons, justice and beneficence. Although this document states individual study proposals should undergo independent review, it recognised the low risk involved with many epidemiological methods and outlined where use of data without consent may be justified.

Surveillance and public health action however, although often adopting epidemiological methods, were not formally addressed in these documents.

4.4.2 Ethical guidelines for surveillance

Surveillance is central to many of the activities undertaken to improve population health, and is performed under legal obligation by all WHO Member States since the 1969 International Health Regulations came into force in 1971[298]. The 1979 Belmont Report highlighted the blurred distinction between medical research and practice, acknowledging they often go hand in hand, though were still able to define each based on the intention

[2] The Belmont report was produced by the American Congress' National Commission for the Protection of Human Subjects of Biomedical and Behavioural Research to provide more robust guiding ethical principles for research following on from the Tuskagee Syphilis Study.

of the activities (to enhance well-being versus to create generalisable knowledge) providing a clear distinction between the two. This distinction means that surveillance-based activities are not subject to ethics committee review, or governed by the principles designed to ensure the ethical conduct of research. Indeed, ethical review may be prohibitive in in surveillance, particularly in emergency situations such as outbreaks, which the CIOMS 1991 report therefore stated should be exempt from ethical review and oversight due to time constraints. Its 2009 revision also explicitly reported the importance of surveillance, even without consent to ensure complete, unbiased datasets, and ensuring the fair distribution of benefits and harms throughout the population. Furthermore, the Nuffield Council on Bioethics, amongst others, suggest it may be necessary to collect even identifiable data for surveillance purposes, that its use without consent may be justifiable and that consent, or an opt-out option may in some situations have serious negative consequences[299, 300].

The 2009 CIOMS Guidelines on Epidemiological Studies however recognised the boundaries between research and practice in the context of epidemiology may be unclear, with many public health agencies using epidemiological methods producing generalisable results as standard public health practice. Despite the clear benefits of surveillance, it is not without risk, largely owing to its infringement of privacy and other civil liberties. These risks, combined with the potential overlap between surveillance practice and research, in the absence of ethics committee review of surveillance activities, have led to the development of ethical guidelines for its conduct, both specific to individual diseases/countries, and more recently, a general international framework from the World Health Organisation[280].

Most ethical guidance has evolved around four basic moral principles; autonomy, beneficence, non-maleficence and justice[301]. Surveillance inherently brings direct conflict between these principals, particularly in respect to autonomy, or respect for persons, with beneficence - with the lack of consent leading to enhanced data and improvements to public health. Ethical justification for surveillance is centred on the balance between these conflicting interests, and it is inevitable that to achieve benefit there may be *acceptable risk* involved, defined as "a risk that has substantially smaller and/or fewer detrimental consequences than the potential hazards of alternative courses of action"[302].

WHO consider four ethical considerations to be key in public health surveillance; *common good*, (beneficence) acknowledging the benefits of surveillance are not to the individual, but shared; *equity*, (justice) provision of resources to those most at need; *respect for persons*, (autonomy), involvement of the public in decision making and protection of data; and *good governance*, accountability and transparency[280]. These considerations underpin their guidance which state countries are obliged to develop appropriate, sustainable, high quality, transparent and secure surveillance systems with a clear plan for analysis and dissemination relevant to public health priorities, minimising harm and taking into consideration the values and concerns of communities. Individuals, they state, are obliged to contribute their data, even without consent when complete data sets are required, owing to the benefits gained to all individuals in the population, including themselves. In addition, states are obliged to develop systems to identify and respond to infectious disease outbreaks and epidemic threats. They state there is a moral duty to use surveillance data to promote health once it has been collected, and it should be used only for legitimate public health purposes. WHO postulate that without appropriate surveillance, countries cannot adequately protect public health or address inequalities, which is in itself an ethical concern, a belief echoed by other groups[303]. These guidelines again address the use of individually identifiable data, the collection of which they argue is justifiable, given its potential to improve accuracy of data, and therefore prevent harm more effectively. Though anonymisation is required for many aspects of surveillance, WHO acknowledges that names, or other personal identifiers may be required for outbreak investigation or contact tracing. However they state that certain uses, for example within vulnerable populations, or when making changes to surveillance systems, may pose complex ethical challenges and require close scrutiny, giving examples such as collecting data that may reveal stigmatising behaviour, or adopting new uses for surveillance data such as for contact tracing. Such changes may require additional ethical oversight, such as through the NHS Clinical Governance Committee, as a UK example.

The use of data for the identification and management of outbreaks is addressed to a degree in WHO's *ethical issues in public health surveillance* guidelines, mostly pertaining to the sharing of data in emergency situations. Though the ethical issues in outbreaks are similar to those in other areas of public health, due to the need for rapid decision making and action often in the context of uncertainty and poor resources, ethical

issues may potentially be overlooked. Drawing from experience following the Ebola outbreak of 2014-16, WHO developed their guidance for *Managing Ethical Issues In Infectious Disease Outbreaks* [304]. These again are based on a set of defined ethical principles. As well as respect for persons, beneficence and justice, which encompasses both equity and procedural justice (transparency, community engagement, accountability and oversight), WHO also highlight the importance of *utility*, where actions are justified if leading to well-being, *liberty*, encompassing a range of political, religious and social freedoms, *reciprocity*, making a proportional return for contributions of others, and, *solidarity* between communities, nations and countries, justifying collective action in the face of common threats. These principles underpin a set of 14 guidelines, addressing obligations to act, community engagement, vulnerable populations, interventions and allocation of resources, restrictions on movement, research and storage of specimens, data sharing and frontline worker's rights and responsibilities.

4.4.3 Literature addressing ethics and surveillance specific to HIV

Owing largely to stigma, surveillance of HIV infection has historically raised controversy and concern amongst individuals[305]. During the early years of the HIV epidemic where no treatment was available, mortality was high and public fear great, exacerbated by the media response to the infection. Groups affected, including homosexual men, immigrants and drug users faced considerable stigma and negativity. At this time, there were legitimate concerns relating to practises such as contact tracing and name-based reporting due to the fears of loss of privacy and further stigma. This led to what has been termed 'HIV exceptionalism'; increased funding and a departure from standard public health practices that favour the rights of the individual over that of the public, in part to avoid driving the epidemic underground[306]. As a result, most early social science research in prevention interventions focussed not on people living with HIV, but 'high risk' members of the non-HIV infected population, due to the risk of further demonising individuals that had been diagnosed[307]. Although with improved knowledge of transmission, and effective treatment from 1996, exceptionalism gave way to 'normalisation'[308], HIV remains stigmatised throughout the world, and transmission behaviours may be viewed by some as morally unacceptable and can in some cases lead to prosecution, meaning HIV requires

additional attention in regards to ethical guidance relating to surveillance and public health practice.

WHO published its *Guiding principles on Ethical issues in HIV surveillance* in 2013[280], and provide a global consensus on the main ethical issues that require consideration in the context of HIV surveillance. They address issues such as name-based reporting of HIV, situations requiring informed consent, surveillance involving vulnerable populations, use of data and dissemination of results. These guidelines acknowledge the risks of 'opt-out' options to the validity of surveillance data owing to the potential for high rates of refusal, however, they give particular importance to the rights of individuals who wish not to participate in surveillance. While they state "it is important to weigh carefully the ethical justifications for allowing individuals to opt-out of surveillance efforts against the public health need for 100% ascertainment of the data", they also recommend, even in non-research public health practice, systems should be created to ensure the principles of informed consent are be respected, developed though consultation with relevant stakeholders. These guidelines also underscore the usefulness of exemption of surveillance data from the law, where a risk of uncovering unlawful behaviours is possible, and state if no such exemption exists researchers have a duty to warn individuals of the limits of confidentiality in this context. Although largely referring to surveillance integrating behavioural and biological studies, WHO recommend that in instances where participants are identified of being at risk, for example due to a lack of knowledge, or access to services, surveillance services should ensure they are referred to appropriate services reinforcing the duty to act to protect health.

Although not formal guidance, Sweeny et al explored the ethical arguments for using 'Data to care' – surveillance data to reengage patients with care or treatment[303]. They conclude that the use of surveillance data for this purpose may be justified if found to be effective, with on-going monitoring and evaluation of risks and benefits. They argue the benefits of engaging, in the context of effective treatment and the reciprocal reduction in onward transmission are high, though the balance with risks of disclosure may vary between subgroups. Benefits include reaching out to those who have never engaged in care, or present to non-HIV departments, who may not be otherwise be identified though alternative approaches for re-engagement, however, on-going evaluation, stakeholder input, operational procedures and legal and policy infrastructure are advised.

4.5 Ethical reflections on HIV phylogenetics

The introduction of molecular typing in infectious disease outbreaks has long raised ethical concerns, largely relating to its provision of more accurate evidence of sources of infection[208, 309, 310], with concerns that issues of blame may deter individuals from engaging with outbreak investigations[209] and raise potential legal questions[210, 311].

With limited real life experience of phylogenetically guided public health interventions, most ethical issues identified in the literature have been through theory and expert opinion, rather than identification of issues raised through practice. Although it has previously been used to investigate outbreaks, its use to guide public health interventions in HIV is a very recent development. In reference to general infectious disease outbreaks, WHO recommends any interventions should be based on available evidence relating to expected risks and benefits, but in the absence of such supportive evidence, it proposes that public health action should be "based on reasoned, substantive arguments and informed by evidence from analogous situations, to the extent possible"[304]. Although TB is arguably less stigmatised[213], molecular epidemiology has been used as standard for many years to help guide public health practice, and in particular to identify and respond to outbreaks, meaning examples of use with TB could theoretically be used as a basis to guide interventions with HIV. However, given the ongoing social stigma surrounding HIV, the fact that it remains incurable, that it is transmitted directly from person to person, often though sexual activity, and that transmission is a prosecutable offense, there are limitations to the conclusions, particularly in regards to ethical issues, that we can draw from experiences with TB.

4.5.1 Exploration of ethical issues in phylogenetic research

Experience in the molecular epidemiology of HIV in the research context has preceded many of the more recent advances in infectious disease genomics, largely due to the fact sequencing of the virus has been a part of routine care for several years, providing a wealth of data for research and public health practice. Although concerns regarding its use in legal cases was highlighted early on[210], in depth exploration of the ethical issues surrounding the use of HIV phylogenetics has gained momentum only in recent years.

Stemming from the recent explosion in phylogenetic research, a group from South Africa aimed to explore the ethical issues surrounding phylogenetics[205], focussing mainly on issues affecting African communities, using the Emanuel Framework[312], a widely used comprehensive framework to guide the ethical conduct of clinical research. In parallel, research within the PANGEA study (Phylogenetics And Networks for Generalised HIV Epidemics in Africa)[313], using whole genome sequencing of the HIV virus to explore transmission dynamics and drivers of the epidemic in sub-Saharan Africa, sparked debate on how this should ethically be used in the research context. Concerns including informed consent and understanding, risks to privacy and the potential to uncover stigmatising, and often locally illegal behaviours (for example homosexuality), led to the development of a review, providing a framework for designing, conducting and disseminating phylogenetic research in an ethical manner[314], summarised from a workshop and on-going discussions between scientists, bioethicists, lawyers, human rights advocates, HIV activists and African community members.

One fundamental previously unexplored area, the perceptions and views of stakeholders, has also been addressed in terms of phylogenetic research in recent years. Mutenherwa et al conducted interviews with experts in the fields of virology, genetics, public health, epidemiology, medical anthropology and ethics to identify the key ethical issues associated with phylogenetic research in HIV to inform future ethics guidelines[315]. Schairer and colleagues similarly performed qualitative interviews with individuals living with, or at risk of HIV and individuals working within the field of HIV (medical and non-medical) exploring the use of HIV phylogenetics in research, to identify the trade-off required between privacy and accuracy to ensure patient and public acceptability of on-going research[207]. These generally found that phylogenetics is deemed acceptable and important given privacy was protected, and risks of disclosure were minimised.

4.5.2 Exploration of ethical issues in phylogenetically guided public health interventions

Although the ethical issues associated with phylogenetic research have been widely considered, the rapid advances in the use of phylogenetics for public health action create new ethical dilemmas. One example is the potential misunderstanding that the data will infer directionality. Although NGS data may be able predict this with reasonable accuracy

in some cases, resistance test data used in this context cannot, though if misinterpreted may cause concerns that it may be used to blame individuals, or even inappropriately used in a legal context. While research requires formal ethical review, informed consent and usually anonymisation of data, public health practice does not. In response to the likely expansion of the use of molecular HIV surveillance to guide public health interventions in the USA, the Third Coast Centre for AIDS Research (TC CFAR, based in Chicago) along with Project Inform held a consultation with participation from US health departments, government funders, researchers, legal scholars, bioethicists and patient advocates to discuss potential benefits, risks and ethical considerations of this use of patient data[316]. The consultation was focused on three key areas, identified through interviews with relevant stakeholders; data security, confidentiality, privacy and potential for misuse; public health action guided by sequence data, and whether consent was required; and benefits and risks, to create an action plan for further research and investigation. A recent review of ethical issues in HIV phylogenetics similarly explored risks and benefits of the use of phylogenetic data in a public health context, concluding further assessments are required to ensure the balance between them is met[317].

4.6 Ethical issues in real time surveillance using phylogenetics

4.6.1 Benefits

The intended benefits of real time phylogenetic surveillance are to identify and disrupt transmission. Real-time surveillance with phylogenetics holds promise in not only providing an up to date picture of the epidemic, identifying dynamic changes as they occur, but also in the potential to identify outbreaks of infection allowing earlier intervention, and provide accurate targets (locations, risk groups) for prevention strategies and case finding interventions. In addition, real time monitoring may provide effectiveness of interventions, though a reduction in incidence, or reduced growth of clusters[179, 318, 319]. Sweeney et al state, when providing a framework for the ethical use of surveillance data for public health interventions, evidence of effectiveness in identifying

the target population and in facilitating their care, and necessity of the use of such data to achieve the intended goals are necessary[303]. Guided by the Emmanuel Framework, the importance of social value and a favourable risk benefit ratio have been emphasised[205], which was also a key theme identified though interview of experts on the use of phylogenetics in research[315]. Accordingly, one of the key recommendations from the TC CFAR consultation was to assess the effectiveness of phylogenetically led interventions compared to standard public health methods[316].

Research that has described public health action based on phylogenetics has suggested benefit, but studies have been small, without control groups[189], and no such interventions have yet been reported in the UK. A previous phylodynamic analyses investigating transmission between MSM within London revealed episodic transmission within distinct clusters, suggesting discrete 'outbreaks' of infection within a densely sampled population arising in the 1990's, with a quarter of transmissions within large clusters occurring within 6 months of infection[183]. Such analyses may suggest potentially identifiable and interruptible transmission chains within a real-time surveillance system, though due to the changing nature of the epidemic, particularly with the introduction of earlier antiretroviral treatment, this may no longer be the case. In addition, the heterosexual epidemic appears to show far less clustering, with fewer early transmissions[320], meaning a cluster-based approach using genetic distances may have little impact. However, outbreaks of HIV have recently been reported in the USA and the UK[180, 321 17, 193, 194], suggesting this may not be the case.

Though these outbreaks provide cases where real-time phylogenetic monitoring would have likely made an impact, updated analyses to characterise current transmission patterns amongst key risk groups are required in order to estimate how real-time surveillance may guide current interventions on a wider scale, and to determine what impact on interrupting transmission they may have. Determining whether phylogenetically guided interventions would be of benefit to the UK population centres on three key areas; the structure of the epidemic; how reliable phylogenetic analyses are in describing the epidemic; and the success of any interventions resulting from phylogenetic findings, areas which require further research and evaluation.

4.6.2 Risks

The major risk of the use of phylogenetic data are social; loss of privacy, autonomy and blame, potentially leading to criminalisation (section 4.5.2.3). Social risk has been defined as 'the danger that an individual will be socially or economically penalised should he or she become identified with an expensive, disfavoured or feared medical condition'[322]

4.6.2.1 Privacy

Unlike current practice in the UK, where phylogenetic data are irreversibly anonymised prior to analysis within the UK RDB, targeting individuals within outbreaks or growing clusters would require a way to relink these data. Arguably, this raises similar issues to name based reporting, as is standard in the US, but not in the UK. Name based reporting itself has raised many concerns, owing to the potential of disclosure[322], though it has been argued that this is a relic of HIV exceptionalism[306] and may be acceptable in an era of effective treatment and prevention strategies. Unlike name-based reporting, phylogenetics has the potential to provide scientific evidence suggestive of sexual behaviours including non-disclosed sexuality[9, 169] and may potentially reveal the identities of transmission partners. This draws parallels with contact tracing, a standard procedure performed with all newly diagnosed patients in the UK, however, with phylogenetics, patients not only lose the autonomy to withhold information they do not wish to share, but, by providing blood for resistance tests, may unknowingly provide investigators with suggestive information on *transmission*, which is not tantamount to the sexual contact identified through contact tracing. Phylogenetic data used in this context can neither confirm directness or direction of transmission, however the misperception that this is possible, particularly given the current advances in inferring directness and directionality in other scientific research, is likely to raise concerns.

Risks to privacy can be mitigated if appropriate measures are taken. Although it is acknowledged that using identifiable data may be a necessity when conducting public health surveillance and intervention, these must be minimised where possible[323], must only be accessible to the appropriate persons, and security measures must be robust[280]. In the UK epidemic, which is much smaller than that of the USA, it is arguable that no

personal identifiers are necessary at all, other than the sexual health clinic, and standard epidemiological and clinical data already collected though national surveillance. Given the reduction in new diagnoses over recent years[256], a growing cluster identified with links to specific regions, or sexual health clinics may provide sufficient information to trigger investigation amongst all recent diagnoses, whilst reducing risk of breaching privacy.

Whilst measures to limit identifiers and keep data secure will reduce risk, the risk of deductive disclosure may still exist[324]. This is the risk that in a specific population, linked demographic data such as race, age range, recency of infection or sexuality may narrow down the possibility of who the sequence data may belong to. In the context of a declining incidence of HIV infections in the UK this risk will be accentuated, particularly in rural regions and amongst minority ethnic groups[325]. This could cause worsening of stigmatisation, marginalisation and blame. However, we can argue that these minority groups may have poorer access to healthcare, and identification though phylogenetics is likely to provide benefit in equalising opportunities for them, particularly in the context of lifesaving treatment and evidence to show earlier diagnosis has both individual and population health benefits. Indeed, it has been argued that using a population based approach using complete surveillance data may help identify those at greatest need who face barriers to accessing care and may not be identified through traditional identification and engagement methods[303]. Therefore, although there is an obligation to share surveillance data[280], in this context, ensuring privacy through appropriate aggregation and censorship when *disseminating* public health findings is crucial.

4.6.2.2 Blame

On a population level, identifiers typically linked to phylogenetic information, such as race or sexuality associated with clusters may could further promote stigmatisation and blame amongst demographic groups if associated with growing clusters and insensitively presented to the public. Again, dissemination of findings needs to be considerate of these threats. Although research findings should be transparently presented, the risks of harm though additional blame and marginalisation of groups may outweigh the benefits in prevention of public knowledge and justify keeping some information out of public knowledge.

The need for individual privacy, as well as reducing the risk of unwanted disclosure of status, is to mitigate the risk of blame. Identification of rapidly growing, or priority clusters suggests higher risk behaviours in the individuals involved, and identifying these individuals as targets for intervention may leave them feeling guilty for these behaviours, accused of transmission, and concerned their sexual partners may be informed they are part of a higher risk network and are likely the source of their infection. Conversely, informing some individuals they are in a cluster, may lead them to believe their partner has been engaging in higher risk activity with others, particularly if they have had few unprotected sexual partners, which may also allow them to deduce the identities of other cluster members. The CDC are explicit that they are not attempting to infer direction[190], which is theoretically possible in some cases with resistance test sequences if appropriate clinical data were available[261], and measures such as having a minimum cluster size for investigation, as was policy in the investigation of a cluster in British Colombia[189], may reduce the risk of inferring transmission between individuals.

Inaccuracies in data may also pose risk of blame. Concordance between phylogenetic networks and contact networks has differed between studies[326, 327], and due to the limitations of phylogenetics, individuals may be incorrectly assigned to clusters they are not associated with, potentially causing issues of mistrust within partnerships, and even mistrust in the healthcare system. However, the limitations of phylogenetics, particularly its inability to determine 'who infected who', may be a protective factor. Weighing up the risk between accuracy of phylogenetic inference and individual privacy is therefore an important consideration. Although practically, the use of NGS data, or more computationally intensive but informative techniques are currently not feasible on a national scale, in part due to the absence of such data for most individuals, but also due to time and cost, advances in phylogenetics mean this may not always be the case. In the future it may be possible to more confidently predict transmission partners with these data, and though this may further enhance public health activity, this needs to be weighed against the potential risk of blame if data are not securely and confidentially handled.

It is also crucial that transmission linkage information relating to others is not fed back to any individual owing to risk of blame. Transmission linkage studies have previously considered this risk, and justified keeping linkage information confidential from

participants given the risk of harm, including violence that may result from unexpected findings[328]. These studies however recruited couples, meaning they both consented to phylogenetic analyses being performed that may confirm or refute transmission between them. In a surveillance scenario, transmission data may provide information relating to other partnerships and no such mutual consent would exist, meaning confidentiality to each individual involved must be respected and disclosure of other clustered individuals must not be an option.

4.6.2.3 Criminalisation

4.6.2.3.1 Criminalisation of HIV transmission in the UK

Currently, in England and Wales it is possible to prosecute an individual who is believed to have transmitted HIV intentionally or recklessly under sections 18 and 20 respectively of the offenses against the person act, 1861[329]. A similar precedent was established in Scotland, where the charge is "culpable and reckless conduct". Reckless transmission occurs when an individual is aware of their diagnosis, of how HIV is transmitted, fails to disclose or to take precautions against transmission and transmits the virus, although transmission is not a prerequisite in Scotland. A conclusion of reckless transmission can only be made in the presence of supportive factual, scientific and/or medical evidence. Recklessness may be supported by the level of risk posed to the complainant, for example, type and frequency of exposure, though informed consent provides defence to the defendant. Informed consent does not necessarily mean the defendant directly told the complainant of their infection, but can be concluded from any knowledge of risk the complainant had received through third parties prior to exposure, or though indirect information, such as being made aware of medical treatment for infection, or observing clinical signs suggestive of infection (for example genital ulceration in cases involving herpes virus transmission). Ideally, both medical and scientific evidence should support an accusation. Medical evidence may include details from medical notes such as dates of diagnosis and negative tests, patients being informed of their diagnosis, symptoms and factors relating to infectiousness, such as viral load. Scientific evidence includes any results from scientific tests. RITA or STARHS testing are one example, though the

limitations of these in confirming recency of infection on an individual level are acknowledged[330]. Phylogenetic analysis is a recommended practice in HIV transmission cases, though the Crown Prosecution Service (CPS) is clear that it can only confirm the absence of transmission with full confidence[331]. Concerns have been raised with the introduction of NGS, that this data may be able to provide evidence to accurately support transmission[332], however recent research has shown, although suggestive on a population level, this still cannot prove direction or directness on an individual level beyond reasonable doubt[160, 161].

Sequence data used may be classified as scientific (blood is drawn with consent for independent sequencing and analysis), or in the case of refusal, previously obtained sequence data may be obtained through court order to determine similarities in subtype, which can only act as supportive medical evidence. In the event of refusal of fresh samples by the accused, theoretically these data, may be sufficient to persuade a jury that reckless transmission occurred; Section 62(10) of the Police and Criminal Evidence Act 1984 states that where consent to take a sample is refused without good cause the court or jury "may draw such inferences from the refusal as appear proper."[333].

Currently there are policies in place to ensure sequences are not released from the UK drug resistance database for medico-legal purposes. Sequence data however 'belongs' to the requesting clinician, and may be retrieved via court order directly from clinicians or laboratories as supportive medical evidence. This is exemplified a case the early 1990s, following an outbreak of HIV in Glenochil prison, Scotland, investigated through phylogenetic analysis, with the inclusion of infected inmates and one heterosexual female contact who had seroconverted to HIV after the release of her partner from prison[178]. The data used for this analysis was subsequently retrieved from the investigator by court order and used as evidence to successfully convict Stephen Kelly, the individual whose heterosexual partner had seroconverted, for reckless transmission of HIV, for which he received a 5 year prison sentence[334]. Although this case is unique, it demonstrates the potential for misuse of data collected and analysed for public health benefit. The use of real-time phylogenetic surveillance would technically have little impact on this process; linkable sequences may still be retrieved via court order whether used for phylogenetics or not, and analyses performed for public health purposes would be of substandard quality for the requirements of court due to the faster methods adopted to allow for large scale

real-time use and the potential absence of appropriate reference sequences, necessitating independent reanalysis.

4.6.2.3.2 Criminalisation and phylogenetically guided interventions

Patients' awareness of the current legal framework has been shown to be poor.[335, 336] Though clear information is available via the national AIDS trust regarding criminalisation and the use of phylogenetics, the information supplied to patients at diagnosis may be variable and misunderstanding regarding the use of phylogenetics as 'DNA evidence' in court may lead to widespread concerns within relating to potential prosecutions using routinely collected data used for real time surveillance, particularly if it is perceived that we can identify 'who-infected-whom'.

Within the context of a real-time surveillance network, phylogenetic linkages between individuals may not be directly disclosed, though there is enormous potential for those contacted through investigations to deduce who may have transmitted the virus to them. In these circumstances, if aware that they were contacted though a phylogenetically guided system, they may believe there is 'scientific proof' available to prove that a reckless or intentional transmission occurred. Even if safeguards may be applied to prohibit the release of this data for medico-legal purposes, potentially the knowledge that this data is in existence may encourage some individuals to attempt prosecution against a partner, or former partner.

In addition, misinterpretation of information, or lack of understanding of legal duties may lead to the reporting of potential 'criminal transmission' by healthcare workers if phylogenetic data linked to their patients is available to them. If for example, clinician directed interventions are applied to a small group of phylogenetically clustered individuals within one clinic, and there is a suspicion of 'reckless' or 'intentional' transmission between these individuals, healthcare professionals may feel it is their required duty to disclose this information in the public interest.

The General Medical Councils guidelines on confidentiality state that disclosure of confidential information may be justified in the public interest, to protect against risk of serious harm, including serious communicable disease or serious crime[337]. It also advises

this public interest must be very carefully balanced against the patient and public's interest in keeping information confidential, including distress to the patient, or harm including disengagement from care. There is an obligation to disclose information if the public are at high risk, however these instances pertaining to HIV transmission are incredibly rare. One recent exception however was the case R v Daryll Rowe [2017], who was successfully convicted under section 20 for intentional transmission to multiple partners. Although this is an extreme example, the huge publicity surrounding this case may distort the public's views on the ease and likelihood of being prosecuted for HIV transmission, heightening anxiety.

4.6.3 Autonomy and community values

As discussed, consent for public health action is not necessarily required and rarely practicable in public health investigations, particularly regarding outbreaks. There have been arguments however that the increased risks of using phylogenetic surveillance may warrant informed consent, and given the lack of evidence behind the use of phylogenetically guided public health interventions, it could be argued it remains in the grey area between public health practice and research.

Consent would remove risks to autonomy posed by the classification of these strategies as public health practice, and allow individuals to be fully informed of the benefits and risks of using their data. It would likely also result in a lower participation rate which would adversely affect the integrity of the data being analysed, affecting the inferences that can confidently be made of the data, possibly rendering it useless. However, if these strategies are used without consent, concerns amongst the public may lead to a detrimental reduction of HIV testing uptake, leading to further transmissions, counteracting their purpose. Indeed, public health bodies and researchers have raised these concerns[338].

Independent of whether consent is required or not, transparency and community engagement are crucial, and it may be that, in the presence of evidence for public health benefit, community consent might be sufficient. It has been argued that the legal authorisations required to weigh up the risks and benefits of surveillance may even act as a form of public consent[339] , though public engagement and consent are necessary to

ensure the suggested benefits are of social value and deemed proportional to the risk involved. This should not only be through standard channels; community organisations and bodies such as the National AIDS Trust, but through communication with people living with HIV, and those who test for HIV.

Informing the public poses great difficulties. The complex concepts involved with phylogenetics are difficult to explain and to understand, even between individuals with a grasp of science. The potential risks involved are equally difficult to explain, as are the benefits, particularly in the absence of hard evidence supporting these. The perception of risk does not appear to be significantly influenced by the likelihood of the actual risk, rather by the magnitude of the potential negative outcome[322] and it is therefore crucial we determine and address the public's perception of this risk and ensure communication of the risk benefit balance, chance of risk, and measures in place to mitigate risk are carefully and comprehensively conveyed.

4.6.4 Are there alternative methods with less risk?

The purpose of the introduction of phylogenetics to guide public health interventions is to identify clusters that may not be obvious using standard clinical or surveillance procedures. Rather than replacing other effective prevention measure such as testing, early treatment and PrEP, these methods may act as an adjunct to HIV prevention; targeting interventions in this manner may become more important as incidence decreases, in an era of scaled up testing, if targets of reaching zero transmissions are to be reached. As described there are numerous risks involved which need to be weighed up against these perceived benefits and ethically, if other less risky options to effectively reach these objectives are available, they should be explored first[340].

One alternative could be to enhance current partner notification policies. This may include more rigorous and intensive case finding related to new diagnoses and automatic referral to a health advisor for repeat partner notification if a patient has a detectable viral load greater than 200 copies/ml. These methods however remain subject to recall and social desirability bias, and may feel uncomfortable, exacerbate feelings of guilt during the early stages of infection and would likely miss smaller outbreaks, particularly if spanning multiple clinics or geographical locations. As a previous analysis of the Brighton HIV

cohort showed up to 74% of acute infections were phylogenetically unlinked possibly suggesting undiagnosed partners, potentially focussing enhanced interventions solely on this group may reach these objectives without the need to create scientific ‘transmission linkages’, reducing the risk of loss of privacy and misuse of data. However again, these methods would be subject to bias, and estimates of the contribution of recent or undiagnosed infection to transmission vary widely, meaning this may lack rigour if aiming to capture all potential ‘outbreaks’. Mapping of infections by postcode has also been explored to try and identify areas, or GP practices for enhanced interventions (S Cavilla, personal communication), however, postcode, or GP practice tells us nothing of where transmission occurred, or who else may be at risk. These alternative approaches, while potentially being more acceptable to patients, therefore would not adequately meet the objectives proposed by this use of phylogenetics.

4.7 Conclusions

The ethical issues surrounding the use of phylogenetics, and to a lesser degree, phylogenetically guided public health interventions have been explored in the literature. The risks associated with phylogenetically guided surveillance can largely be attributed to a loss of autonomy and risks to privacy, although much theory is deduced from very little real-life experience. Although these risks are present with many types of disease surveillance, the lasting stigma associated with HIV and possibility of criminalisation mean enhanced measures are required to maximise security and enhance acceptability of this use of data. Given a current paucity of evidence for its benefit, we are currently unable to determine whether the benefits outweigh the potential risks discussed. With on-going research in America, and potentially from work in Glasgow, this evidence may soon appear, providing a case for similar interventions throughout the UK.

Alongside the question of whether this may provide benefit in the UK, several important questions remain; how we can communicate the risks and benefits of phylogenetics to the public; whether informed consent is necessary, and what people will and will not consent to; and whether protection from judicial use can and should be applied to these data. Even if the issues raised in this chapter are addressed, real life experience will likely clarify the relative importance and consequences of these, and raise new issues, meaning exploration

of the ethical issues associated with phylogenetically guided interventions requires on-going review and reassessment in this rapidly growing field.

Chapter 5: What is the acceptability of using phylogenetic data in clinical and public health practice? A qualitative study aimed at intervention development.

Chapter purpose and summary

As described in chapter 4, various guidelines, studies and ethical reflections have provided insight into the ethical issues surrounding the use of phylogenetics in HIV. One key area that has not yet been explored are stakeholder's views of how, or even whether, these interventions may be conducted in a manner that is acceptable and minimises harm to the individuals involved. The aim of this chapter therefore is to explore views of people living with HIV, people that test for HIV and healthcare workers who may be involved in phylogenetically guided interventions to determine how acceptable interventions need to be developed, what harms may ensue from unacceptable use, and safeguards required to mitigate these harms.

5.1 Abstract

Background:

The use of phylogenetics to guide HIV public health interventions (PHI) is a growing area of interest, and has potential applications in the UK, with recent evidence its implementation may have interrupted a large outbreak in Glasgow. There has been little exploration into the acceptability of this data usage amongst patients or healthcare workers. A qualitative study to map acceptability was conducted, determining:

- how phylogenetic data can be used acceptably in the context of HIV PHI
- what safeguards are required
- negative outcomes that may result from unacceptable use.

Methods:

Focus groups and interviews were conducted with people living with HIV, HIV negative MSM and clinical staff in two major UK cities. Illustrated explanations of phylogenetic concepts and clinical vignettes describing potential uses of data to guide PHI were presented. Audio recordings were transcribed verbatim and analysed thematically.

Results:

Although views differed both within and between demographic groups, the use of phylogenetics for PHI was generally considered acceptable, though with significant caveats. We identified three underpinning themes; stigma; blame and prosecution; and public health responsibility. Acceptability was determined by the balance of perceived personal risk (stigma or blame) vs. public benefits. Acceptability was highest when the potential prevention effects were greatest, and/or data were used anonymously. Despite the use of identifiable data pushing this balance towards personal risk, we identified several factors that were found to make even de-anonymised analyses more acceptable. These included:

- restriction of access to identifiable data to the direct healthcare team
- confidence in robust security measures
- protection of data from use in criminalisation cases
- informed consent
- increased understanding of the limitations of phylogenetics
- drawing parallels with standard uses of data for prevention purposes.

In the absence of adequate protections or understanding, many participants felt the use of phylogenetics for PHI risks discouraging HIV testing and engagement in care.

Conclusions:
When implementing phylogenetics to guide PHI, patient understanding and confidence must be ensured in order to address concerns and avoid disengagement from HIV testing or care. Although we identified key factors influencing acceptability within this geographically limited sample, expanding our intervention alongside the development of phylogenetic approaches would be advisable given the ongoing exploration of novel strategies and potential for widespread use.

5.2 Introduction

Recently, the use of real-time sequencing and phylogenetic analysis of baseline HIV drug resistance tests as a public health tool has been expanded throughout the USA, identifying phylogenetic clusters of public health concern, allowing targeted interventions including re-engagement into care, ART adherence support and partner services in an attempt to prevent further transmission and detect any undiagnosed cases associated with the cluster[204]. Although exploration of the ethical issues and community engagement has been highlighted as a priority within the evaluation of these interventions[190], acceptability of these actions has not yet been adequately explored.

Previous work focussing on HIV in the USA has shown molecular epidemiological studies to be acceptable to patients, with the benefits being seen as outweighing personal risk, although threats to privacy were a concern. Although examples of its use to guide public health interventions were provided to illustrate the power of molecular epidemiology within this study, topic guides were based on its use in research, and interviewees were asked specifically to respond giving their views on its use in this context, rather than for public health purposes[207]. Published research describing the use of molecular epidemiology to aid case finding identified through systematic review reported few patient related barriers or facilitators, and as such, the acceptability of phylogenetics for this purpose is still unknown (see chapter 2). Although the use of

general surveillance data, including CD4+ T cell counts, to guide public health interventions has been found to be acceptable, using phylogenetic data to detect transmission and potentially trace contacts raises a number of potential concerns and emotional reactions. These include issues around prosecution[209, 210], alongside loss of privacy[208] and blame[209], which have not yet been fully explored in the context of public health interventions. Although during a public debate held to determine perceptions of whole-genome sequencing in Tuberculosis, over half of participants believed that individuals who may have acquired TB from a presumed source should be able to access their information for medico-legal purposes[341], perceptions of such use in HIV, which is arguably more stigmatised infection HIV and more commonly associated with criminalisation, are likely to differ significantly.

The use of phylogenetic data may be ethically justifiable if able to uncover undiagnosed infection and prevent onward spread, however harm may result, for example reluctance to test, failure of disclosure of potentially infected contacts or refusal of resistance testing, if community understanding and consent is not gained, or if adequate safeguards are not applied. For research with public health interventions using phylogenetic data to be carried out, particularly with potentially improved discrimination of transmission events with newer sequencing technologies such as next generation sequencing (NGS) [342, 343, 344], it is crucial we understand patients' concerns, how these may be addressed, and what level of use of phylogenetic data in public health interventions is acceptable.

By determining patients understanding, concerns and/or misconceptions, and potential outcomes of the public health use of such data, future HIV phylogenetic research will be informed as to how phylogenetics may be used, or what level of integration of personal data or identifiability may be acceptable on a larger scale for public health interventions, and how research may be designed to minimise the risk of harms to any individual. This will also provide guidance on what levels of consent and information would need to be provided to patients if using next generation whole genome sequencing for routine care.

5.3 Research objectives

5.3.1 Primary research objective

- To determine the acceptability of the use of phylogenetic data for public health interventions in HIV in the UK

5.3.2 Secondary research objectives

- Determine patients' understanding of the use of phylogenetics in HIV, including its use as evidence within criminal transmission investigations
- Determine what patient concerns might be relating to the use of their sequence data for public health interventions
- Determine any potential adverse outcomes (for example, reluctance to test for HIV) of the introduction of routine use of phylogenetic data for public health purposes
- Determine whether and how any patient concerns may be addressed to allow the use of their data in this context
- To explore the role and nature of informed consent in this context
- Determine how the introduction of NGS may alter these perspectives

5.4 Methods

5.4.1 Qualitative design

Owing to the research questions asked of this study, which aimed to develop knowledge of practical use in the development of an intervention, a pragmatic approach was adopted as it is not committed to any single epistemological or ontological stance. Rather, a pragmatic approach aims to identify the most appropriate methods for the problem at hand. The pragmatic paradigm develops practical solutions to real life problems whilst

avoiding philosophical debate and thus fixed, often abstract principles, acknowledging there is not one final truth, as with other apporaches[345]. Rather than truth being objective, and independent of the mind, as with critical realism for example, or being entirely a social construction, as with constructivism, pragmatists believe truth is actively created through practical application of experience and thus ever changing. Pragmatism aims not to achieve absolute certainty but to adopt methods most suitable to answering the research question to address issues in the real world and formulate effective solutions to problems[346]. It acknowledges that nature is unpredictable and every situation is uncertain justifying research flexibility and often mixed methods approaches[347], and focuses on human behaviour and experience rather than abstractions to improve problematic situations in a diverse and complex world.

As the aim of this study was to explore the views of theoretical interventions to identify issues with acceptability, focus groups and interviews with key stakeholders were selected as the most appropriate method of data collection. Focus groups were selected to elicit a wide range of views, while interviews allowed in depth exploration of individual issues, and thematic analysis was identified as the most appropriate analytic method[348]. This variation in data collection allows for richer and more complex data generation, than the single method approach.

Thematic analysis allows researchers to identify patterns, or themes within rich datasets. It allows flexibility in terms of epistemological approaches, and as well as being a factor within other qualitative methods, has emerged as an accepted method in its own right, with recognised methodological steps to ensure rigour[348].

Other methods were considered but deemed inappropriate; grounded theory owing to the fact the aim was not to develop theory, but to explore the views of individuals in regards to potential interventions, ethnography due to the fact it primarily aims to provide rich description of populations, and narrative analysis as a broad range of views, rather than only in-depth individual accounts, was required. Co-design[349], or co-creation[350], an emerging method used frequently in design, were considered promising alternative methods for data colection. Within these, participants would help create phylogenetically guided interventions with the research team providing insight into what might be acceptable. Given however both a lack of evidence for which phylogenetically guided

methods may be most effective in terms of prevention, and a lack of understanding of public and patient views and comprehension of the subject matter, individual and group discussions focussed around multiple potential uses of phylogenetics with explanation from facilitators in the event of misunderstanding, analysed to identify underlying themes was felt to be an important first step in order to obtain practically useful knowledge for development of this newly emerging field.

5.4.2 Participants and recruitment

Recruitment was conducted between December 2017 and May 2018. The majority of participants were recruited from the Brighton Sexual Health and HIV clinics. This is a mostly white, MSM population, in a city with the highest prevalence of HIV outside London. In addition it is a well-informed population with high uptake of participation in research, including previous phylogenetic studies[88, 351] making it a potentially suitable population for piloting phylogenetically led interventions. To gain representation from non-MSM groups, recruitment was also conducted via community organisations within Brighton and London to target a more ethnically and gender diverse HIV positive population. Participants were recruited according to specific self-reported demographics for interview and for each focus group; HIV-positive MSM diagnosed within 5 years, HIV-positive MSM diagnosed more than 5 years ago, black African men, black African women, HIV negative MSM, and healthcare workers within the field of HIV.

Participants fitting the target demographics of HIV positive focus groups were directly approached at Brighton's HIV clinic, and in addition flyers were distributed via UK community groups (The Sussex Beacon and UK CAB). Those approached in clinic that were interested provided contact details to arrange focus groups discussions, and those who received flyers contacted the researcher directly. Individuals not wishing to, or unable to attend a focus group were offered one-on-one interviews. HIV-negative MSM were recruited though a dedicated MSM sexual health clinic held weekly in Brighton's Claude Nicol Sexual Health centre. Potential participants provided contact details to the researcher, or took a flyer containing the researchers contact details and participant information leaflet if they were interested in taking part or wanted more information. Patient participants were offered £20 as a 'thank you' for taking part. All participants

were informed the primary researcher was a clinical doctor working in HIV undertaking research towards a PhD in phylogenetics for public health.

Healthcare workers were informed of the study during a weekly all-clinic meeting and asked to contact the researcher if they were interested in participating. Due to difficulties in arranging an evening group, the group was conducted during the weekly academic morning where no clinical work takes place.

5.4.3 Procedure of interviews and focus groups

Focus groups were held according to the demographics of the participants. Most were held in a quiet waiting room at the Elton John Centre for clinical trials in the early evening, when there were no patients or staff in the department, and two were held at HIV-community group venues. Focus groups were facilitated by the primary researcher (myself, a female with no practical qualitative experience) and a male Research Fellow with extensive experience in qualitative research (Alex Pollard). After informed consent was received, patient-participants completed a short demographic questionnaire dependent on their HIV status (Appendix F). Healthcare workers' clinical roles were recorded, but no other demographic or personal information was collected owing to the small number of individuals working within the department and therefore risk of deductive disclosure.

Introductory questions assessed participants' knowledge of use of personal data for audit, research and surveillance purposes and understanding of anonymisation. To gauge understanding of phylogenetics, brief introductory questions followed by illustrated explanations of resistance testing, genetic sequences and phylogenetics were provided, allowing participants time to provide their reflections, and ask any questions or clarify misunderstanding. Vignettes were then provided, describing current and potential future uses of phylogenetics, with specific vignettes focussing on specific uses of data or key issues identified from published literature, conference abstracts and discussion with phylogeneticists, HIV community representatives and individuals with specialist knowledge in law and HIV criminalisation (table 1). Vignettes were chosen to elicit responses to potential real life scenarios and identify factors that may influence

acceptability, allowing participants to explore possibilities surrounding very sensitive topics. Each vignette was intended to explore and exemplify a specific ethical issue pre-identified by the research team. However they were designed to be open to exploration of issues that were not expected. Due to the complexity of the information and concepts being introduced, most explanations and scenarios were presented with simplified, or 'cartoon' images to illustrate and make them easier to understand. These were presented as PowerPoint slides though a projector as the topics or vignettes changed (Appendix G).

Following the discussions, a short debrief was provided, clarifying how their data is currently may and may not be used, and reiterating that these scenarios discussed uses of phylogenetics that were not performed in the UK. Participants were given the opportunity to directly discuss any issues or concerns with the researchers that the discussions may have raised. They were also provided with a small handout explaining phylogenetics, what resistance test sequences may be used for and are not used for in the UK, and reassurance that criminalisation is rare with a breakdown of what criteria are required for transmission to be potentially considered reckless (Appendix H). The contact details of the researcher were provided within this in case of any issues related to the discussions that participants did not feel comfortable raising on the day.

Table 10: Vignettes and pre-identified issues

Scenario	Issues raised	Groups	Reference
1. An analysis of UK sequences reveals groups of heterosexual men cluster with groups of MSM, suggesting some heterosexual men may not disclose MSM behaviours to Health-care workers	**Attitudes to the use of anonymised phylogenetic data as per standard practice**	**HIV+FGD, HIV-FGD, HCW, IDI**	**Hue et al (2011), Ragonnet-Cronin et al (2018)**
2. All sequences in the UK, including you own, are added to a phylogenetic network in real-time. A large cluster of infections from PWID in Brighton is identified and public health action is targeted at this group. The data are linked only to aggregated epidemiological data.	**Attitudes to the use of anonymised data in real-time**	**HIV+FGD, HIV-FGD, HCW, IDI**	**Ragonnet-Cronin et al (2018)**
3. The public health team identify a concerning cluster and are able to identify individuals to whom the sequences in the cluster were taken from. The clinic contact you, as your sequence is within this cluster, to ensure you are on/adherent to ARVs and ask about any recent sexual contacts that might need testing.	**Attitudes to the de-anonymisation of data within a phylogenetic network**	**HIV+FGD, HIV-FGD, HCW, IDI**	**Poon et al (2016)**
4. The identities of your sexual contacts (HIV positive and negative) are added to this network to help guide interventions including allocation of PrEP.	**Attitudes to the addition of sexual contact data to a phylogenetic network**	**HIV+FGD, HIV-FGD, IDI**	
5. A friend has heard about phylogenetics for public health interventions, and wants more information before testing so comes to you for advice.	**Impact of the use of phylogenetic surveillance HIV testing**	**HIV+FGD, HIV-FGD, IDI**	

Scenario	Issues raised	Groups	Reference
6. A friend thinks their partner transmitted to them recklessly. They want to prosecute him and ask your opinion of this, and what information the police might need.	**How the use of phylogenetics for criminalisation of HIV influences attitudes**	**HIV+FGD, HIV-FGD, IDI**	
7. The public health team are highly concerned about a rapidly growing outbreak of HIV, and are using your phylogenetic data, potentially linked to some details about you without your consent.	**The need for informed consent**	**HIV+FGD, HIV-FGD, HCW, IDI**	
8. A new way of processing resistance tests has been introduced for routine clinical care that looks at more of the HIV gene. It's a little quicker and cheaper for the NHS. If it is used for phylogenetics links between sequences may be more accurate.	**Acceptability of and concerns regarding NGS**	**HIV+FGD, HIV-FGD, IDI**	
9. Your patient who recently disengaged and stopped treatment has been found to cluster with several recent infections. One new patient names them as a sexual contact.	**Deductive disclosure of a potential transmission source**	**HCW**	
10. A newly diagnosed patient thinks they were sexually assaulted, and acquired HIV as a result, but have poor recollection of the events. The police closed the investigation after no leads were found. Their vague description reminds you of another recent diagnosis. They are both found to be part of a new cluster of infections you are investigating.	**Being aware of circumstantial evidence collected for public health purposes and obligations towards sharing knowledge**	**HCW**	

5.4.4 Data analysis

Interviews and focus groups were digitally recorded and transcribed verbatim, and digital recordings were listened to and transcripts read and reread to ensure familiarisation with the data.

Preliminary annotations and notes were made of initial thoughts and transcripts were imported into Nvivo for coding. Initial coding was performed on the entire dataset, with all data items considered and coded irrespective of their immediate relevance at this time. Initial codes were then collated into potential themes. Although vignettes were designed to raise specific issues, due to a lack of previous research on this topic at the time it was conducted, a largely inductive analytic approach was used to identify patterns and themes within the dataset. These themes were reviewed, initially by re-reviewing coded items within each to ensure they were coherent and agreed with the meaning of the theme. Secondly, all themes were reviewed against the research questions posed and further refined, and finally the dataset was recoded against these themes to ensure all relevant data were captured. A doctoral student with experience in qualitative analysis read transcripts and reviewed coded items to validate the identified themes.

5.4.4 My role in the study

The study concept and aims were designed by myself, along with my supervisors Professors Jackie Cassell and Andrew Leigh Brown, and Dr Jaime Vera. The study design was conceived by me, Jackie Cassell and Professor Bobbie Farsides. I created a protocol, participant information and consent forms, flyers and detailed interview schedules with PowerPoint sides with input and feedback from the above team, and from members of community organisations and experts in HIV and criminalisation (Mr Matthew Williams, Mr Edwin Bernard and Dr Robert James) and gained ethical approval. I recruited most participants from clinics, who were initially approached by members of the clinic team, and others were recruited through flyers distributed in clinic or via community groups. I conducted all interviews and led focus groups, with assistance of Alex Pollard. I checked

and amended transcripts and performed coding to identify themes, which were validated by Shanu Sadwani, and discussed with Jackie Cassell and Alex Pollard.

5.4.5 Reflexivity

Prior to my doctoral work, I had worked as a physician for seven years, specialising in HIV and sexual health for the latter three, with the final year based in Brighton. During my doctoral studies I continued to do a weekly HIV clinic and weekly on call shifts and maintained good relationships, on both a professional and personal level with most members of staff. My dual role within this department required consideration prior to commencement of fieldwork for two reasons; the effect it would have on me as a researcher, both internally and in the eyes of participants, and on my views of the use of phylogenetics for public health interventions.

Staff (and my previous 'regular' patients) were aware of my current position as a research fellow; a predominantly research role with some clinical commitments, and owing to these commitments, I was viewed first and foremost as a doctor. A positive aspect of being familiar with the staff and surroundings was my ability to blend into, and understand the setting. However, this also had the potential to cause difficulty, as my integration relies on the fact that I have an on-going clinical role within the team, and as such, I may have been called upon for clinical duties while recruiting, or be engaged in conversation regarding administrative tasks, or even 'friendly chat', all of which will may have diminished my role as a researcher. To be effective as a researcher and be able to focus on this aspect of my current role required temporary detachment from clinical duties. Separating my roles however posed potential, though unlikely, ethical challenges that required consideration as my duty as a doctor to those in need must always be prioritised. Fortunately, despite most recruitment and interviews being performed in a clinical setting, the chance of a medical emergency requiring my input was small, and the chance of there being no other immediate medical cover smaller, though these situations were planned for. One additional personal barrier secondary to my dual role was how I perceived my colleagues regarding me as a researcher. Being a senior registrar, I am confident in my clinical practice and I feel my colleagues consider me as someone who is knowledgeable in the field, and as such approach me for advice or patient review as needed. As a researcher, I had essentially reverted back to 'student' status, and despite

this being at a doctoral level, I had concerns they may underestimate my skills and therefore the validity of my work. Although did not affect the quality of my research, I had to make a conscious effort to present myself with confidence to avoid this during recruitment in clinic, and during the healthcare worker focus froup.

Having been focussing my research on phylogenetics for the prior two years, in combination with my previous (and still in part on-going) role as an HIV physician, I had preconceived ideas around what participants may feel about phylogenetics, however my own views on the subject were a little conflicting. A positive view of the use of phylogenetic data for PHI would be advantageous for me as a researcher, as being given 'the go ahead' from stakeholders might allow me to continue work in this direct field; however this is where my conflict mainly lies. As a clinician working in HIV, I am aware of the heightened importance of privacy, having encountered situations where accidental disclosure of a patient's status to a GP has led to formal complaints and huge emotional turmoil, and as an advocate for my patients I am very wary of any interventions that may put privacy at risk, and potentially increase stigma. I, like many other HIV physicians and advocates believe criminalisation of HIV transmission promotes fear and stigma, and the potential of phylogenetic data for PHI to be used in this context raises personal concerns. Having explored the ethical issues and potential harms such use of data can bring, and anticipating the emotional anxieties this could cause patients, I was therefore almost a little fearful of real-time phylogenetics moving ahead without thorough consideration of these issues. This internal conflict I felt was actually positive, as not having a firm position on whether the use of phylogenetics for PHI is acceptable myself likely reduced bias during discussions.

The blurring of boundaries between healthcare worker and qualitative researcher have been discussed in the literature[352], and potential strategies to overcome the associated difficulties vary. Addressing these potential barriers to my work required a personal approach. To separate my roles I made an effort to divide time between each role; I used an out of office for my clinical email address and checked and responded to these at planned times. I informed the clinic staff when I was present for recruitment or interview purposes that I would not be taking part in any clinical work unless there was an emergency, and there were no other persons present that were qualified to deal with it. One further step I took was to modify my appearance to distinguish my two roles, by

dressing in a way that was smart, yet ‘unsuitable’ for clinical work by disregarding hygiene measures that are necessary to practice as a doctor by not tying back my hair, wearing jewellery and wearing long sleeved tops. Adopting a consistent, yet different, ‘research uniform’, and wearing my university ID, was an easy way for my colleagues to distinguish which role I was in.

5.5 Results

5.5.1 Participants

40 patient participants were recruited, with a further 9 health care worker participants. One patient participant did not contribute within a focus group discussion so was excluded from the analysis. Self-reported characteristics are shown in Table 2. The median age of patient participants was 36, and all but one reported sexual risk for HIV. Of the 9 healthcare workers, five were doctors (consultants and specialist registrars), two were senior HIV nurses and two were health advisers. It was not possible to calculate the number of participants who declined to take part in the study owing to the fact we were unable to quantify the number that picked up or saw a flyer.

Table 11: Self-reported characteristics of non-healthcare worker participants

Parti-cipant	Sex	Age range	Ethnicity	Sexual-ity	Risk factor for HIV	Education	FGD/ IDI*
1	Male	50-54	Missing	MSM	Sexual	School	FGD 1
2	Male	25-29	Missing	MSM	Sexual	Graduate	FGD 1
3	Male	45-49	White British	MSM	Sexual	Diploma	FGD 1
4	Male	45-49	White other	MSM	Sexual	Graduate	FGD 1
5	Male	30-34	White other	MSM	Sexual	Graduate	FGD 1
6	Male	40-44	White	MSM	Sexual	Graduate	FGD 2
7	Male	30-34	White British	MSM	Sexual	Diploma	FGD 2
8	Male	40-44	White British	MSM	Sexual	School	FGD 2
9	Male	40-44	White British	MSM	Sexual	Vocational qualification	FGD 2
10	Male	50-54	White	MSM	Sexual	School	FGD 2
11	Male	45-49	White	MSM	Sexual	Graduate	FGD 2
12	Trans male	25-29	Missing	MSM	Sexual	School	FGD 2
13	Male	45-49	African	Het	Other	Vocational qualification	FGD 3
14	Male		Missing	MSM	Sexual	Missing	FGD 3
15	Male	50-54	Missing	Het	Sexual**	Graduate	FGD 3
16	Male		Missing	Het	Sexual**	School	FGD 3
17	Male	45-49	African	MSM	Sexual	School	FGD 3
18	Male	40-44	African	Het	Sexual**	Postgraduate	FGD 3
19	Female	55-59	Black African	Het	Sexual**	Vocational qualification	FGD 4
20	Female	50-54	African	Het	Sexual	Missing	FGD 4
21	Female	45-49	African	Het	Sexual	Missing	FGD 4
22	Female	45-49	African	Het	Sexual	Graduate	FGD 4
23	Female	50-54	Missing	Het	Sexual	Vocational qualification	FGD 4
24	Female	Missing	Missing	Het	Sexual	School	FGD 4
25	Male	50-54	White	MSM	Sexual	Graduate	FGD 5
26	Male	45-49	Missing	MSM	Sexual	Diploma	FGD 5
27	Male	40-44	White	MSM	Sexual	Graduate	FGD 5
28	Male	40-44	Missing	MSM	Sexual	Diploma	FGD 5
29	Male	Missing	Missing	MSM	Sexual	Missing	FGD 5
30	Male	Missing	Missing	MSM	Sexual	Missing	FGD 5
31	Male	Missing	Missing	MSM	Sexual	Missing	IDI
32	Male	Missing	Missing	MSM	Missing	Missing	IDI

Parti-cipant	Sex	Age range	Ethnici-ty	Sexual-ity	Risk factor for HIV	Education	FGD/ IDI*
33	Male	40-44	White British	MSM	Sexual	Postgraduate	IDI
34	Male	45-49	Black	MSM	Sexual	Postgraduate	IDI
35	Male	45-49	African	Het	Sexual**	Graduate	IDI
36	Male	40-44	Black African	Het	Sexual	Diploma	IDI
37	Female	35-39	Black African	Het	Sexual	Diploma	IDI
38	Female	45-49	African	Het	Sexual	Diploma	IDI
39	Male	Missing	African	MSM	Sexual	Graduate	IDI

Het = heterosexual *FGD = Focus group discussion, IDI = In-depth interview. FGD 1: MSM diagnosed > 5 years ago, FGD 2: MSM diagnosed < 5 years ago, FGD 3: Black African men, FGD 4: Black African women, FGD 5: HIV negative MSM

**Also reported blood transfusion as a risk factor

5.5.2 Analysis

5.5.2.1 Patient perceptions

Most participants had little prior knowledge of phylogenetics, were unaware this may have been performed using their data, and detailed explanation was required to achieve understanding. A wide range of views and opinions were identified through the data, and although data was collected according to specific situations, the findings were underpinned by three major themes; stigma, blame and prosecution, and public health benefit, with acceptability dependent on the balance between these factors. We also identified four further themes which directly influenced the acceptability of any scenarios, as discussed below.

Table 12: Emergent themes

Underpinning theme	Subthemes
1. Stigma	1. Stigma of HIV compared to other medical conditions 2. Lack of knowledge perpetuating stigma 3. Previous experiences of stigma
2. Blame	1. Blame and judgement 2. Prosecution
3. Health benefits and responsibilities	1. Public health benefits 2. Personal benefits
Theme	**Subtheme**
4. Data management	1. Access to data 2. Data security 3. Third party use
5. Autonomy	1. Transparency 2. Informed consent 3. Consent of contacts 3. Logistical issues and patient understanding
6. Accuracy of data	1. Limitations of phylogenetics 2. Misinterpretation of data 3. Next-generation sequencing
7. Parallels to standard practice	1. Parallels to current standard practice 2. Maintenance of current standards of care 3. Concerns relating to standard procedures irrespective of phylogenetics

Underpinning theme 1: Stigma

Despite participants acknowledging stigma and fear surrounding HIV had reduced since the introduction of effective treatment, the impact of stigma played a key role in how acceptable the use of phylogenetics might be, owing largely to the potential impact of stigma if their status was disclosed through such analyses or investigations.

> *'I think, I think mostly there may be some people might object to it because it's embarrassment, I don't think, I think they're scared or embarrassed to tell the other person because of the stigma they see to it, but otherwise if there wasn't a stigma I think people would be quite happy with it, they would not …'*
>
> (Black African man, IDI)

Subsequently, several participants suggested some HIV negative individuals may be reluctant to test for HIV due to fear of disclosure owing to the associated stigma, if they were aware phylogenetics may be performed on their sequences and not fully informed of its purpose and safeguards.

> *'…this phylogenetic, you're going to be making references to so many cases that will cause a lot of self-stigma, that will cause a lot of deflection from clinical case, from taking medication, from the wellbeing of the people that you would be targeting, that's my question. It's not going to be something that will cause a stoppage for some clients from accessing support from the hospital, or [unclear] it will cause something like stigma within the environment.'*
>
> (Black African male, FGD)

'HIV is different to other diseases'

The use of phylogenetic clustering was felt to be more sensitive in HIV than it might be in other conditions, as HIV was considered to be different to other illnesses, mainly due to it being currently incurable, requiring lifelong management and associated with stigma. Although one participant did raise parallels with phylogenetic surveillance and identifying 'clusters of cancers' another retorted 'But we're not talking about cancer', suggesting sensitivities around HIV revolved mostly around the stigma relating to it, rather than the potential severity of illness.

> *'But I just think it's just a different, it's a different level. It's like you know, the stuff like you know, more common things that happen, and are easily treated. HIV is something that needs to be treated for life.'*
>
> (MSM diagnosed more than 5 years ago, FGD)

Ignorance breeding stigma

Several participants reflected on experiences with individuals who lacked knowledge of HIV and therefore treated them in a stigmatising way, reinforcing their desire to keep their diagnosis confidential. This was particularly apparent most black African groups, who discussed being isolated or treated differently by GPs or hospital staff, enhancing their wariness of disclosing their status even to medical professionals outside of their direct HIV team.

> *'The stigma will never go away, I'm telling you. Never in life, people are not willing to come down to it, even if it you think about the doctors, the GPs who think they are doctors, but if they are not specialised in this, it is very tough for them, they always say no. You go to them, no. They still have that kind of stigma because they are not aware of it.'*
>
> (Black African male, FGD)

Personal experiences

Though the impact of stigma was apparent throughout all groups, it emerged particularly strongly amongst black African men and women, many recollecting experiences from the earlier days of the epidemic from their home countries. The impact of these experiences necessitated strict confidentiality to be maintained. Amongst these individuals, concerns about stigma resulting from disclosure, particularly within close-knit communities was a major issue.

> *'…in 1987, when we come from Nigeria when the use of HIV common, they call it AIDS there in Nigeria, the picture, the very first picture was given to us on the TV, on the national TV, so somebody that's skinny! People see that mentality up til now, til this this moment, that people diagnosed, maybe they are just in their mind, because they first reference that people had, they store it in their memory, they store it in their memory and they know that analogy, they can all be dead, up til now, people still believe that who living with HIV can be skinny and have all sort of grudges.'*
>
> (Black African male, FGD)

Underpinning theme 2: Blame

Blame and judgement

Concerns surrounding blame for transmission were apparent with both anonymised phylogenetic data, and with the introduction of potentially identifiable information. Primary concerns relating to anonymous data were of potential blame of specific groups for causing the epidemic if findings were not sensitively disseminated. Many referred to

stereotypical views that their demographic group were the 'cause' of HIV and were concerned results of analyses may perpetuate blame, and hatred towards them.

> *'And it's drug users that are passing it on, it's black men, Asian women, whatever. And then it's like, you know, because they say this is literally, we're not saying this could be happening, we're saying this is happening. And then it's like, you know, if they start to get targeted, because it still is, you know, the stigma is still there. And also for any, you know, for any hate crime, you know, they just see a target and they do it, you know. That would be my only concern.'*
>
> (MSM diagnosed more than 5 years ago, FGD)

When presented with scenarios depicting small local outbreaks detected through anonymised real-time analyses, some participants raised concerns relating to deductive disclosure, or false assumptions of identity from linked demographic data, particularly amongst smaller, or minority groups, suggesting despite anonymisation, there may be a risk of individual harm depending on how results are presented.

> *'Yeah, things like they, you know, I think people don't know how long it's been, oh they moved out from London to Bristol at that time, it's them, you know. They're a black straight man, it's them. It's just like you know, people get it into their heads in the wrong way sometimes.'*
>
> (MSM diagnosed more than 5 years ago, FGD)

> *'If I'm thinking Isle of Wight, I'm thinking slightly different kettle of fish to that or I'm thinking of closed communities, rural communities and then local level data being held around potential outbreaks in potential areas. I think the ability to be able to, dare I say, point the finger.'*
>
> (MSM diagnosed within 5 years, IDI)

Deductive disclosure was not a great concern for most, who were reassured by the fact that *their name* was not attached to the data, meaning they felt they could not be identified with certainty. This primacy of not using a name, which was more apparent in the black African groups may however reflect a lack of understanding of the possibility of deductive disclosure, or faith that 'anonymised data' means this is entirely untraceable.

> *'Um, I don't feel nothing because there's no, my name is not there.'*
>
> (Black African female, FGD)

> *'Use the information, but take out my name.'*
>
> (Black African male, FGD)

When discussing the use of identifiable data, concerns were raised that clinicians may judge them for partaking in high-risk behaviours or blame them for transmission, which they suggested may put some individuals off testing or engaging with care.

> *'...clients will not come for the clinic again 'cause when you go the clinic, and your assessments say, "you sleep, you sleep, you sleep…"'*
>
> (Black African male, FGD)

The issue of blame for transmission not only related to being blamed, but blaming others for transmitting to them. Some participants acknowledged they blamed a specific partner when they were diagnosed, and others acknowledged they did not know who they had acquired it from. Those who were unaware stated that although at the time, they were angry and wanted to know who they had acquired it from, most now accepted they wouldn't ever know and it was better that way. Discussion of phylogenetics brought forward a possibility of finding out this information which caused discomfort, and it was generally felt that this knowledge should not be made available to patients, due to a concern that it may encourage anger or reignite issues they had learnt to deal with.

> *'Well I'm just trying, I don't know who I contracted HIV from, so if then that was to escalate into the consultant or the GP having to tell the police and then I was linked to that and then it might be a way or means of me finding out who passed on the virus to me, and how am I going to feel when I do actually, 'cause I actually like the fact that I don't really know now. 'Cause when I got first diagnosed it raised 101 questions…'*
>
> (MSM diagnosed within 5 years, FGD)

Primarily however, there was a fear of being blamed for transmission if data were identifiable, which was apparent throughout all groups. There were concerns that results may be misinterpreted and used to wrongly blame individuals of transmitting intentionally. It was strongly felt that data should not be able to identify 'transmission links' between pairs of individuals and identities should not be shared with members of the same clusters, several participants stating that fear of being blamed may deter people from testing.

> *'It can discourage a lot of people, because what people don't want is to be blamed to say 'you're the one that gave it to me'. Because most people are in denial and, you know, even if they know that they're HIV and they pass it on so someone, they don't want the blame. So they'll rather just it go quietly.'*
>
> (Black African female, FGD)

Prosecution

The biggest concern relating to blame was the risk of prosecution. Although all participants were aware of criminalisation of HIV, they had mixed understanding of the intricacies of laws surrounding HIV transmission. Although most understood 'reckless transmission' is a prosecutable offense, several sought clarification as to whether having an undetectable viral load through treatment definitely means the virus is untransmissable, providing protection from the law, and whether condoms and disclosure of status are always required by law if having sex with a serodiscordant partner irrespective of condom use and 'being undetectable'. Although this was not explicitly stated, a lack of clarity of behaviours that may put one at risk of prosecution, or knowledge of factors that may be protective, may have increased concerns about the use of phylogenetics if it is perceived that practices or behaviours they have experienced may put themselves at risk of prosecution. Most participants were aware that transmitting when undiagnosed is not classified as reckless transmission, though it was acknowledged others may not be aware of this, and may be reluctant to test in the light of phylogenetic surveillance in case they had unknowingly transmitted to someone else.

> *'Is it by law that I have to tell every girl I sleep with that I've got HIV?'*
>
> (Black African male, FGD)

> *'Um I've been told, I don't know how accurate the information is, but I've been told that if you've got a undetectable viral load that you can't pass HIV on as a rule of thumb. So where would you say legally on that?'*
>
> (MSM diagnosed more than 5 years ago, FGD)

Most participants, particularly MSM, felt strongly that in general it takes 'two to tango', with both parties being equally culpable if transmission occurs, arguing people should take responsibility for their own status and not blame, or attempt to prosecute others.

> *'I think a lot of people, if you go round having unprotected sex then you've got to expect that you could get something and the blame is always yours anyway. You know, you can't blame other people…'*
>
> (MSM diagnosed more than 5 years ago, IDI)

In addition, there was scepticism as to whether there was any point in police involvement even if warranted, as 'you can't prove anything', with some feeling the police were underqualified to manage transmission cases, or simply 'too busy'. As such, there were major concerns that data collected with good intention may be used in prosecution cases,

which heavily influenced how acceptable the use of phylogenetic data might be. Although generally, the black African groups were more accepting of prosecution if behaviour was considered genuinely 'reckless', they were still largely opposed to police involvement, one of their primary concerns being that prosecution cases may expose your HIV status to the public.

> *'Yeah, it breaks the confidentiality, because the media will come in, they want to know, you know, it will be in the newspapers and everything. […] And if they decide they want to go to the police, yes the police can in the end solve it. But after, it's a long battle, it's not something that can be done overnight. It's going to take years, and expose you as well.'*
>
> (Black African female, FGD)

Although several participants stated the use of phylogenetics performed for public health purposes in court would not concern them personally as they didn't think they could be implicated in a transmission case, a major concern was that other innocent people may be wrongly prosecuted using phylogenetics if incorrectly accused of behaving 'recklessly', which they felt may influence HIV testing uptake. Almost all participants did however recognise that very rarely prosecution for intentional transmission with the use of phylogenetics was justifiable, several referring back to the recent case where Darryl Rowe was convicted of intentionally transmitting HIV to several partners. They also acknowledged the ability of phylogenetics to refute transmission made it a useful tool. It was still felt however data collected for public health purposes should not be used for prosecution, and any phylogenetics performed for this purpose should be done so independently.

A small number of participants however, saw the potential for prosecution using phylogenetics as a positive aspect, suggesting if people were aware phylogenetic testing could prove if they had transmitted HIV, they would modify their behaviour, and take more precautions against transmission.

> *'But what I want is the research to help stop spread and … yeah, this will help because what stops spread. Because once the word goes out that now there is a thing, which can actually say that it's you. I'm telling you. You'll think twice, even if you're drunk, you will use a condom.'*
>
> (Black African male, IDI)

Underpinning theme 3: Health benefits and responsibilities

The third major theme was personal and public health benefits responsibilities.

Public health benefits

Although benefits proposed within vignettes included personal benefits (potential earlier access to treatment, modification of behaviour, re-engagement into care) primarily participants were very accepting of phylogenetic use if it could improve outcomes for other individuals, including earlier diagnosis and prevention of further transmission. This was despite there being little direct personal benefit to themselves, most stating they were engaged in care and on suppressive treatment.

> *'It would be a little bit embarrassing, for us it's sort of, but within you, especially for the younger generation you'll be happy to do whatever it takes to eradicate or reduce it. So you'd be happy to do it.'*
>
> (Black African woman, FGD)

Participants had a very altruistic response when they felt their data may have public health or prevention benefits and generally were more and more accepting towards even fully de-anonymised analyses as the possible benefits increased. This was exemplified by the introduction of a 'large uncontrolled outbreak' scenario, when after several participants had previously raised concerns relating to the use of de-anonymised data, most were happy for any identifiable data to be used, even without consent if this would assist with its control.

> *I get that they have to have people's names to make sure if they're going to contain it, so yeah again I'd be fine with that.*
>
> (MSM diagnosed within 5 years, FGD)

Some participants commented on how without research involving people living with HIV in the past, HIV would not be the manageable condition it was today, and felt that in the light of benefits to others it was their responsibility to contribute their data for phylogenetics for the benefit of the next generation.

> *... but overall I'm saying that the benefits of this kind of study outweighs the risks of whatever you call, the negatives. 'Cause I remember HIV as… as an issue, has come a long way. If people who are not involved, including those who have passed on, in the earlier studies and researches we would probably not be where we are today, so I think it is incumbent upon us also to support these kind of ideas, if it is to benefit both us today and those who are in the future.*
>
> (Black African male, FGD)

One HIV negative participant however, who found the concept of phylogenetic monitoring too intrusive felt that these suggested benefits would not be not great enough when compared to standard HIV prevention interventions to support the use of phylogenetics. One recently diagnosed MSM participant who underwent a one-on-one interview also reflected on human behaviour and disagreed with the sentiment that people would be happy to dismiss their privacy to help others, branding people as essentially selfish and narcissistic. Although this disagreed with the general findings, potentially others may have internally shared these views, but didn't feel comfortable stating them in front of others at a risk of sounding 'self-centred'.

> *'I think as human beings we like to think that we're ultimately altruistic and that we're selfless and I don't feel that is necessarily an appropriate way to think of humans and human behaviour underpinning the reasons why we would tackle it in this way. People are not necessarily ultimately altruistic particularly in modern day society, people are ultimately highly selfish, very egocentric, narcissism, you know when we look at how society's changing many, many people like to present that they are there for the benefit of others, but ultimately as part of the human psyche it's self-preservation and looking after their self-first is core to people's sense of identity and I think that for me is the massive issue that this would simply not tackle.'*
>
> (MSM diagnosed within 5 years, IDI)

Personal health responsibilities

Another positive factor was personal responsibility for health. Participants prioritised their health and as such, didn't feel the use of phylogenetics would impact on their testing behaviours or engagement in care. Several explained they either tested regularly, or tested because they were symptomatic or took risks, and all agreed it was a priority to get tested and start treatment if positive. Many did acknowledge however, there were individuals who do not share this sense of personal responsibility, though suggested these people are unlikely to test irrespective of whether phylogenetics was being performed or not.

> *It's just like you know, I don't think you know, they may say oh you'll have phylogenetic testing, you know, when you are tested. And people are like still got to go and get, I've got to go and get tested.*
>
> (MSM diagnosed more than 5 years ago, FGD)

Balance between underpinning themes influencing acceptability

Overall, the use of phylogenetics to guide prevention strategies was acceptable, though with major caveats. This largely centred on the balance between individual harm (disclosure or loss of privacy leading to stigma, blame or prosecution) and public health or personal benefit, and therefore was dependent on the situation presented and on the individuals perception of risk and sense of personal responsibility to others. Personal factors that appeared to influence this balance included being on treatment or in stable long-term relationships, mitigating risk of prosecution and blame, though attitudes varied both within and between demographic groups.

Generally, anonymous analyses, even performed in real-time were acceptable, due to low perceived risks to the individual and potential public health and research benefits. It was felt however, sensitivity was required in the presentation and dissemination of results to avoid any further stigmatisation of groups, and analyses that risk deductive disclosure should not be shared publicly.

For some individuals moral responsibility outweighed any potential risks posed to them, even with fully identifiable data, suggesting *any* use of their data was acceptable if there would be a public health benefit.

> *'it's worth your privacy being invaded to that extent that you've protected someone else. We have to put a stop to it at some point and say well all reasonable measures have to be used.'*
>
> (HIV negative MSM, FGD)

> *'But moreover it's just like, you know, the more important thing is stopping the spread of it if he is the cause of it. It's just like you know, were that to be wrong though, it's just like who knows. But you know, I think it overrides anything just, you know, protecting the public health.'*
>
> (MSM diagnosed more than 5 years ago, FGD)

For most, however, de-anonymisation increased the possibility of disclosure, pushing the balance in favour of personal risk. Several participants, when faced with an initial scenario involving identification of individuals linked to a phylogenetic network, immediately stated that it felt 'too intrusive', however several factors were identified which were felt to reduce this risk, increasing acceptability for almost all participants. Despite overall acceptability of the use of phylogenetics if the risk benefit balance was

met, there were two individuals (one recently diagnosed MSM, and one HIV negative MSM) who appeared to be strongly opposed to any use of phylogenetics that may be in anyway de-anonymised, irrespective of safeguards. One being sceptical of the potential benefits of phylogenetically led interventions over current prevention efforts, believing the invasion of privacy was too great for minimal public health impact, and one explicitly stating human nature is to always protect oneself over the protection of others.

> *'I think we might be, we might be over estimating what the issue is. I think that people that absolutely don't want to go into a clinic or will never set their foot in their clinic are a minority in society. I think most people are well informed, most people get tested when it's time to get tested when there's a concern that their health might be at risk or they might have been exposed. I think most people who have sex with other people regardless, no, that's not true (laughs), I think most men that have sex with men will go to the clinic to get tested even if there's no reason for concern.'*
>
> (HIV negative MSM, FGD)

> *'For my own benefit … it would be very helpful. Whether I would actually care about the benefit of others is a very different question to ask me.'*
>
> (MSM diagnosed within 5 years, IDI)

Additional factors that influenced acceptability included management and security of data, consent and transparency, awareness of the limitations of phylogenetics and drawing parallels to current surveillance practices and HIV related interventions. Again, the influence of these factors varied between individuals, with subtle differences between demographic groups.

Theme 4: Data management

Access to identifiable data

Access and sharing of data very much centred on confidentiality. Although participants were comfortable with researchers handling their anonymised data, they felt identifiable information should only be handled on a 'need to know basis'.

> *'It all, for me personally it boils down to who has access. If it's restricted to the relevant people having the right access for the right reasons, no problem. It's when it becomes open source, when anybody can gain access to it, or if it gets*

> *leaked to the wrong person. That's when I have an issue. Because it's information that is personal, and it can have a devastating effect on people's lives if the wrong people get hold of that information'.*
>
> (MSM diagnosed > 5 years ago, FGD)

Most participants were happy for their direct HIV team to see and deal with de-anonymised phylogenetic linkage data, but were very reluctant if anyone else could access personally identifiable data, and the handling of this data by only the direct HIV clinical team was one of the most important factors in enhancing the acceptability of de-anonymised phylogenetics. They also felt anonymised analyses where there was a risk of deductive disclosure, should not be shared with anyone outside the direct healthcare team, most importantly not with the public. In part, this comfort with doctors handling this sensitive information was due to the fact that individuals stated they disclosed information about sexual partners and risk behaviours to their doctor anyway, and several declared they trusted their doctors, particularly those of black African ethnicity, with many referring to identifiable data that is only accessed by the healthcare team as still being 'anonymous' and 'confidential'.

> *'So such information I think, it is the relationship you build with your clinician that they can actually extract most of this information, the majority of it, and so long as it is used in the context of the system, to address something I think it wouldn't bother anybody. What is sensitive to them…'*
>
> (Black African male, FGD)

> *'Presumably though this data, although you know, it can identify you, but it's still held anonymously. So only the clinic would be able to tell who you actually are then. So even if somebody from the public did have this information you would still remain anonymous. It doesn't identify you individually.'*
>
> (MSM diagnosed > 5 years ago, FGD)

Concerns particularly related to police access, which almost all participants were opposed to, with several stating it would affect their testing behaviour if information collected for public health purposes was used for prosecution. Although a few felt they should be able to access relevant data if appropriate, they felt this should be through a doctor, with information shared on a strictly need to know basis. This main issue with police access was again that it may be used to convict innocent individuals, and that attempts at prosecution may expose their HIV status to the public, but also expanding access to the police raised concerns that data may be available to other public bodies and even the public, compromising the security of the information.

> *'No, no, 'cause once you have too many people accessing that information it is liable to be, it's likely to be lost along the way. It's likely to be abused, you know, you need to contain it, you need to have one body looking after that information and only that body can be approached, 'cause once the cops have that information then they can, anything can happen, we are humans and can make mistakes and pass it on to somebody else, and then it can leak, before you know it, it's all over.'*
>
> (Black African male, IDI)

Data security

Participants frequently made reference to data breaches within the NHS including cybercrime, incompletely wiped computers, loss of USBs and accidental sharing of confidential data via email or post, and consequently some appeared sceptical that the level of security required to make the use of identifiable data linked to phylogenetics acceptable could be achieved. Most participants felt that phylogenetic data should not be directly linked to any identifiers but be linked to random codes, the key for which should only be accessible to the direct healthcare team.

> *'How that may impact on people's propensity or willingness to want to engage with treatment if they wear the anxious/cynical hat about what happens to that information. I also then question that, if as individuals what we know about ourselves in terms of the human psyche, we can't always trust ourselves one hundred percent, how can that be instilled by a public body if there are historical issues relating to breaches of data and confidentiality.'*
>
> (MSM diagnosed within 5 years, IDI)

Third party use

Lastly, patients were highly opposed to their data being used for any purposes other than public health, seeing this as 'misuse' or 'abuse'. Several individuals raised concerns their data would be used 'to make money', some raising concerns about selling of data to third parties including pharmaceutical and insurance companies, which may at some point be used against them in some way. They consequently required much reassurance that their stored data would not be used for non-health purposes in the future.

> *'Yeah, and it's just like you know, it's okay at the time when this information is stored and you were told exactly what's going to happen to it. That then changes, you know, so it's like you know, it could be that initially it's like it will only be medical professionals. And then in five years' time it's like well actually we've sold it to, you know, this medical company so they can, you know, and things change.'*
>
> (MSM diagnosed > 5 years ago, FGD)

> *'...how far does, does that information get used and can it be, you know would it then be eventually used for things like prosecutions or say, for example, I don't know, say we ever got private health care or insurance, you know could insurance*

companies then access that kind of information and they'd go well you participated in this risky behaviour, blah, blah, so we're not going to insure you, etc, etc.'

(MSM diagnosed within 5 years, FGD)

Theme 5: Autonomy

Transparency

There were differing views on when consent was required, but generally participants felt that the use of 'low risk', anonymised data did not require consent. Although participants were mostly unsurprised and unconcerned their anonymised HIV sequence and demographic data may be used for phylogenetic analysis, many did state there should be more transparency as to how their data may be used, though there was a divide as to whether this was something that was 'nice to know' or something they 'needed to know'. Others felt informing patients about phylogenetics as currently used was unnecessary, in part as there was no risk involved or direct effects on the individual, and because patients might either not understand, or become suspicious that their data may not be fully anonymised, or used for purposes other than research or health care.

'But, we should just be told that, you know, your information that we hold can be, you know, sometimes used by researchers for trials and stuff like that, so be aware. And then at least we know where we stand. So yeah, that way would feel comfortable, but how they use it? You know, as long as it's for the good of eradicating the virus, it's good.'

(Black African female, FGD)

'I think if they was to keep informing us of everything that they did behind the side, in the background, they'd never get anything done'.

(MSM diagnosed > 5 years ago, FGD)

Informed consent

Consent was felt to be important where de-anonymisation may occur, due to the increased risk to the individual, with some feeling very strongly about this**.**

'If it was anonymous I wouldn't care so much, you know, it doesn't matter. If it's just anonymous data and it's just got like my age and you know, my general demographics, then that's fine, you know, it can't be linked back to me, but if it was some… If it had various specific information or detailed information about

> *me attached then yeah, I would want a consent in that situation. I would be really, really annoyed if it wasn't.'*
>
> (MSM diagnosed within 5 years, FGD)

Participants equated to the use of personal data without consent as a loss of autonomy, or personal freedom, feeling that they were being monitored against their will or revealing issues they may not want to disclose.

> '*Where do you lose the choice to choose'*
>
> (HIV negative MSM, FGD)

Several raised concerns, for example, relating to heterosexual men who have had sex with men and not disclosed this information, and whether the use of phylogenetics linking them to MSM clusters is a breach of privacy, and may further 'push them in the closet'.

> *'Or, do we continue down the path with, "Okay, so I can't get the information from you out of your mouth, but I can get the information from you, from other influences in your life – zip code, bloodwork and certain habits, you know.'*
>
> (MSM diagnosed > 5 years ago, IDI)

Although some felt consent was not necessary for use of identifiable data if the information was only securely held by the direct healthcare team and used for prevention purposes, many participants felt fully informed consent was required irrespective of safeguards and purposes, and an opt out option should be offered. Most participants said they would opt in if safeguards were applied, though acknowledged there were others who would not want their data used in any circumstances, and felt their choice should be respected. They also felt if data was used in scenarios that patients may feel uncomfortable about without consent, this may damage the doctor-patient relationship, and lead to reluctance in sharing information with their medical team in the future.

This need for consent however was recanted by most when discussing an urgent outbreak situation requiring public health action, where the public health benefits of its use were seen as a priority over personal risk, with most happy for any identifiable information to be used if it could assist with preventing further spread. A small number however still felt informed consent would be required for use of identifiable data in any scenario, particularly amongst black African groups for whom disclosure of status was a major

concern, where despite reassurances they expressed concerns that their data may be released to the public in this situation.

> *'Yeah, the next thing when they see my name and my address, they just Google my body to see me now.'*
>
> (Black African Female, FGD)

Conversely, a small number felt that consent should not be required when using phylogenetics for public health purposes, raising concerns that incomplete sampling may lead to incomplete investigation and a consequent reduction in public health benefit, with one suggesting consent could be received covertly without appropriate information provision as there would be greater benefits to all.

> *'This might sound horrible if we don't know at all…we wouldn't have any concerns, just printed on the form that we didn't bother reading, that said oh you need to sign it here. Yeah, I'm still on it, that if people don't bother reading things, and they don't bother, the majority of people don't bother to opt out and it's not going to hurt anybody, it's only going to benefit people.'*
>
> (HIV negative MSM, FGD)

Consent of contacts.

When faced with a situation where sexual contacts, positive or negative, were linked to a network to aid outbreak investigations, some participants felt explicit consent was required from their contacts to disclose their names or any personal details. Others however drew parallels to standard care, where contact details, including name, testing outcomes and sometimes contact details are stored in patient records. Nevertheless, the idea of loss of autonomy for use of data was an ongoing concern.

> *'For me there's a consent issue there. It's one thing that if those people I'm connected to consent to be in this thing and for me to pass on their information I guess, but it's kind of another if they don't. I'm just sort of going, here you go clinic! And the clinic goes, alright we're going to use this…'*
>
> (MSM diagnosed < 5 years ago, FGD)

> *'Yeah, because what will happen is when I go for HIV test they will ask me, about um, do you know where you got it from? And you're not sure, so you might say oh I've been seeing this man, and he's the only person that I've been seeing for such a long time. And they ask you are you happy for us to contact him? And, and I'll say, I will talk to him first. So I will go to him and tell him look, I've had an HIV test and mine came positive, so I think you should go and have yours tested. And then hear what he says, then you can now tell your doctor who he is. But if he tells you no, no, I can't tell the doctor because that's um breach of, you know, confidentiality. So it's up to him. I can tell him, and if he says no then…'*
>
> (Black African Female, FGD)

Logistical issues surrounding transparency and informed consent

Despite differing views on if and when consent or provision of information was required, the majority agreed there would be many challenges involved, and no real consensus regarding how and when to consent was reached.

Timing

Participants stated that when consent was required, it should be received at the point the data was being collected and used, i.e. shortly after diagnosis. However, many acknowledged this was a difficult time emotionally, and at this point most participants had very little knowledge of HIV. It was felt that discussing phylogenetics at the point of diagnosis would add an additional element of complexity at a time where there was already so much to process, and as a result, they felt this information would either completely 'go over their head', or cause anxiety, due to the fact they had not yet built trust in the system. Two participants discussed the process of partner notification being an additional pressure at diagnosis. Though they now understood the purposes behind it, it may suggest that a lack of awareness of standard procedures at this time could make the idea of phylogenetically led interventions appear even more intrusive. Many therefore felt the discussion of phylogenetics and receipt of consent should wait until an individual has come to terms with their diagnosis, which should be determined on a case-by-case basis by their physician.

> *'I think the early stages are, of diagnosis, yes there's a lot to process, there's a lot to process about the self, there's a lot to process about health. I think to just add one element about the whole consent regime is a very small issue to think about compared to the bigger issues …and I think if you are able to dedicate time to think about the bigger issues around management, treatment and all of that, for me early stages that's when you want the greatest level of security and safety and resilience to be developed …and therefore I think the early stages to hold those discussions sooner rather than later for my experience …'*
>
> (MSM diagnosed < 5 years ago, IDI)

Conversely, one participant felt the point of diagnosis was a good time to discuss phylogenetics *because* the myriad issues they were facing meant there would be less overthinking and therefore fewer concerns. Others that agreed with this notion however felt that it should be re-discussed at a later point, as in the event of patients forgetting, or

not absorbing the information they may lose trust in their clinical team if their data is used in the future.

Almost all participants felt the use of phylogenetics should not be discussed prior to testing positive. Partly as a negative test result would render this discussion pointless, but most importantly this was because it would add a further layer of anxiety to the testing process, which may deter some individuals, who may think their data might be used against them.

> *'I think there's no sense in saying it beforehand anyway because if your test comes back negative then it's a moot point so you may as well only tell somebody once they get a positive diagnosis, so… It's a bit of a waste of time otherwise, again because I think it would put people off testing '*
>
> (MSM diagnosed < 5 years ago, FGD)

> *'It's so frightening testing isn't it, and it's so stigmatised still and I think there mustn't be anything in place that further dissuades people, erm, and I think that […] might dissuade people…'*
>
> (MSM diagnosed < 5 years ago, FGD)

Understanding and complexity of information

A further complication of the consent process was the complexity of the concepts and issues surrounding the use of phylogenetics. The majority of participants had never heard of phylogenetics, and although during the discussions participants developed understanding of the issues discussed, they required careful explanation with images to illustrate key points, and frequently misinterpreted information, requiring clarification. Common misconceptions were that phylogenetics would be used to definitively prove one person transmitted to another, despite explanation that it was not possible to determine who-infected-who with certainty and reassurance potential investigations would not be undertaken in a way that would identify direct transmission. This may suggest that patient concerns may be so great that despite reassurance they may automatically 'fear the worst', and dispelling these anxieties and gaining trust in the system may take a great deal of explanation and counselling. Within these discussions, care was taken to ensure participant understanding and a significant amount of time was spent explaining concepts and addressing misconceptions or misunderstanding, which is not always possible when explaining procedures within busy clinical settings. Participants therefore understood that consent would be a lengthy process, particularly early in

diagnosis, where general knowledge of HIV was poor, and emphasised the importance of educating patients at an early stage.

Communication

Due to the complexity of information that required explanation, and possible misinterpretation and anxiety that could be caused, several participants felt that taking consent would very much rely on how the information was communicated. It was felt that although communication needed to be transparent, it was important to emphasise the public health intentions and safeguards applied in order to alleviate any concerns. Some also suggested that input from trained counsellors or psychologists may be beneficial in helping address patient concerns relating to these processes.

> '*This is about the wider clinical agencies or bodies, so you are referring to psychological services here […] when I look at the level of psychological services to meet people's emotional needs there's an imbalance there, so the question is what would be the resource made available for people to be able to access to give that balance …'*
>
> (MSM diagnosed < 5 years ago, IDI)

Some participants suggested information about phylogenetics should be distributed, or 'advertised' to the general public, to inform of the potential benefits of such a system and gain public support. This idea however, contradicts the opinion that individuals should not be made aware of phylogenetics prior to testing positive, unless it was carried out in a way to fully inform every individual that may consider testing for HIV, based on the principle that a 'little bit' of information, rather than receiving none, or developing a full understanding of the facts, may heighten unease.

Theme 6: Accuracy of data

Limitations of phylogenetics

Several participants were very much reassured by the fact these analyses cannot definitively say who infected who and were opposed to anything that may point to individual transmission. This provided comfort in terms of risk of prosecution; although it was accepted this may be used as supportive evidence, it could not definitively conclude

transmission from one person to the other. Conversely, some participants felt accuracy would ideally be improved when using phylogenetics for public health interventions, in order to enhance investigations and maximise public health benefits, whilst reducing the chance of patients being falsely implicated within outbreaks.

> *'…purely because you're not going to be able to tell which way it was transmitted. If they've transmitted to them, or have they got it transmitted? I don't think that would necessarily put anybody off testing. I wouldn't imagine so.'*
>
> (MSM diagnosed > 5 years ago, FGD)

Misinterpretation of data

Participants however also saw the limitations of phylogenetics as a negative, owing to the risks of misinterpretation within analyses. This manifested as concerns that analyses may suggest transmission from specific groups, incorrectly laying blame people of a certain demographic for transmitting the virus. When discussing investigations involving de-anonymisation, participants also raised concerns that misinterpretation may incorrectly associate them with a cluster of infections. It was suggested such false associations may provoke feelings of self-stigma and self-doubt. Participants also raised concerns that if they were in a stable relationship and were contacted for being associated with an outbreak, it may cause mistrust within their relationship despite awareness of the fact that phylogenetic linkages do not equate to transmission between the individuals whose viruses are linked.

Next-generation techniques

When the concept of next-generation sequencing, with more discriminate phylogenetic results was introduced, interestingly participants were generally unconcerned, many seeing this increased accuracy as a positive aspect if it facilitated public health investigation. Though some, on initial description appeared alarmed that this might reveal 'who-infected-who', most were reassured that it cannot confirm direct transmission with complete certainty. In addition, as previous scenarios had addressed safeguards required for acceptability, participants interpreted its use to be in the context of these safeguards; data only used for public health purposes, not used to falsely incriminate, kept secure and only accessed by appropriate people, which made its use, even for public health intervention, acceptable.

> '*So somebody who say didn't know they had HIV and to then get them on treatment so then they became undetectable etc, so for that purpose, absolutely,*

that's a good thing, […], as long as it was primarily used for that purpose and not other creepy reasons.'

(MSM diagnosed < 5 years ago, FGD)

Theme 7: Parallels to standard practice

Parallels to current practice

Although few patients were explicitly aware that their general HIV data is used anonymously for surveillance purposes, most presumed this was the case and appeared unconcerned that this was happening. Participants frequently acknowledged that they regularly provide information regarding sexual partners if testing is required, and several stated their doctor knows 'everything about them' anyway. Therefore, even in de-anonymised analyses, given identifiable information is only accessible by the direct healthcare team and identities of members of clusters are not shared with other members, they drew parallels to current care which made this use of their data appear more acceptable. Several participants felt that phylogenetic surveillance, particularly if anonymous, is similar to a level of health monitoring they would anticipate routinely, some stating, somewhat apathetically, that it is probably already occurring or will occur in the near future anyway.

'They're doing it anyway because if you're going to be having sex and you come to a clinic anywhere in the world, like this, you're being monitored because you've got a piece of paper with your name and everything written on it, it's no different, this is now just a code.'

(MSM diagnosed < 5 years ago, FGD)

Participants also acknowledged, in many cases individuals will be aware, or strongly suspect they know who they acquired their infection from, or if there was a risk they may have transmitted to someone else, irrespective of phylogenetic analysis. Some stated that this information would come to light anyway, so adding this type of monitoring would make little difference to this aspect of the investigation.

'…they probably have got a good idea anyway, in a small place like this.'

(MSM diagnosed < 5 years ago, FGD)

Maintenance of standards of care

Participants were also reassured that current standards of practice would be upheld throughout the scenarios, which was found to be reassuring. The most important factor was the duty of confidentiality doctors have to their patients, given participants largely only found it acceptable for the direct healthcare team to access identifiable data, and had major concerns regarding potential leakage of information. This was particularly apparent when individuals raised concerns about their identities being shared with sexual partners, often rebutted by other participants stating that the healthcare team would still maintain this duty to the patient, as with current partner notification practices.

> *'…they wouldn't tell you anyway because it would surely be confidential, they will keep that information confidential and if they weren't able to track you down to a specific individual they wouldn't go around telling everybody else involved because, well patient, doctor confidentiality, they wouldn't be allowed to, so we'd never know, even if your consultant did…'*
>
> (MSM diagnosed < 5 years ago, FGD)

Concerns relating to standard procedures irrespective of phylogenetics

Throughout the discussions, many participant concerns did not relate to phylogenetics directly, but to interventions that take place irrespectively including contact tracing. Many participants did not seem to be fully aware that these were in fact standard care despite having likely experienced them at some point in the past. Issues included the option of sending anonymous texts to contacts of infection, and in a few instances, providing the details of sexual contacts to their healthcare team for this purpose. These issues frequently dominated discussions over the issues directly related to phylogenetics, and suggest that not having full understanding of current practices or not being informed of the precise differences phylogenetic surveillance would make may influence acceptability.

Healthcare workers perceptions

Concordance with patient perspectives

Public health benefits

Healthcare workers generally took a patient centred approach to the issues surrounding the use of phylogenetics, and as such, themes raised corresponded with those of the patient participants. Again, the overall discussion centred around risk to patients (primarily of prosecution) vs. the potential benefits of using phylogenetics for public health purposes.

> *'But it's like anything, data can be used positive but also, alternatively it can also be used very negatively against someone. It's working out how would you avoid the negativity and can you completely mitigate against it being used negatively or what can you do to try and safeguard the negative use of that data, that's really the key thing I think. And I don't know what the answer is.'*

Healthcare workers recognised potential benefits of interventions led by phylogenetic surveillance, giving the example of the recent outbreak amongst PWID in Glasgow and suggesting there may be other smaller scale outbreaks happening 'under the radar' where phylogenetically led interventions may provide benefit. In addition, it was seen as a potential way of prioritising investigation for individuals who may have disengaged from care, particularly in the context of earlier and more effective treatment meaning these individuals may become a more important transmission source in the future. It was however also acknowledged that this particular clinic had very robust measures in place for assistance with reengagement into care, and the benefit in terms of this aspect may be limited locally, though it may be of more use in other centres with lower retention rates.

Blame

Healthcare workers were however concerned about blame and criminalisation, which made added a level of caution to the discussion. They felt that although phylogenetics cannot confirm transmission with certainty, and clinicians may already make assumptions as to whether a patient may have transmitted the virus to others, it added an element of confidence in these assumptions, which felt to them 'like blame'. They acknowledged the stigma that remains around HIV, and the fact that managing it requires more sensitivity that other curable, or less stigmatised infections, and were concerned that

implicating patients within outbreaks may enhance self-stigma, and impact negatively on their mental health.

> *'It does feel like blame for an already stigmatised condition and I think, as well we can't be 100% sure of what we're saying to that patient either. We're again making a presumption which we've already made anyway but we're just not saying it, whereas if we have a phylogenetic test that is telling us something and we're imparting that information again it's a presumption, but we're showing a patient a black and white result on a piece of paper potentially, that will make them feel, as you say, responsible for other people's infection when they didn't knowingly what to transmit to anybody.'*

Data management

Healthcare workers strongly agreed that only they should be able to access any data that is identifiable, and lead investigations, reporting only anonymised outcomes back to public health bodies. They acknowledged the sensitivities in investigating potential transmission of HIV between individuals and felt if this was required, only those trained specifically to manage such situations should be involved, to reduce potential negative impacts on individuals if not managed carefully. They agreed robust security measures needed to be place given the sensitive nature of the information, and potential risks if data were leaked, and also agreed data should not be used on a level which may suggest direct transmission between individuals.

Autonomy

Healthcare workers also raised the issue of consent. Participants highlighted that there are patients that have very strict views about how their data can be used and shared, and this use of data, even for surveillance would be highly unacceptable. In addition, the use of resistance test data was likened to research studies where consent was required for additional use of their anonymised DNA from blood samples, however, generally it was alluded to that they felt the use of completely anonymous phylogenetics did not require consent. For the use of potentially identifiable data however, it was felt that unless we were able to remove risk (mainly the risk of prosecution) consent would be necessary and a detailed statement, including what the data can and cannot be used for would be required from the body overseeing the programme, to ensure all risks are understood.

Impact of phylogenetics on clinical care and clinician responsibilities

When presented with scenarios where there may be a risk of deductive disclosure of potential contacts or transmission sources, HCWs were generally unconcerned owing to the fact that this occurs in usual practice through partner notification, offering examples from their practice, and confidentiality measures in place meaning this posed little risk to any patients. They did however acknowledge there were some issues in regard to how they needed to act on this additional data. If for example, data suggested one of their patients may have transmitted HIV during a recent outbreak, should they inform them of this fact? On one hand, informing them may lead to self-stigma and blame, however would not informing them directly of this fact inadequately address their behaviours, leading to further transmission? Similar issues relating to their responsibilities arose when they suspected a patient may have committed sexual assault by knowing they may be linked to the victim from information provided in the investigation of a recent small phylogenetically identified outbreak, for which the police had insufficient evidence to identify the attacker. Though such a scenario would be a rare occurrence, the fact that HCW were unsure how they should act in light of this 'circumstantial evidence' caused great unease.

Judgement

It was highlighted how a lack of judgement towards individuals is crucial when working within the field of HIV.

> *'But lack of judgement is crucial really, to us keeping engaged in, in keeping people engaged in care, because I think if people feel judged they run.'*

One healthcare provider raised concern that access to data associating their patient with an outbreak scenario may provoke unconscious internal judgement. This caused discomfort, as they felt it may impact on the way they cared for their patient.

> *'Because you're a human being you think well why did you do that? You know, you're, you know, you're potentially, you know that you're not on treatment, you know that you've got X, Y… You know, why are you having unprotected sex with all these people, you know that you're exposing them to risk. Erm, you know I mean I do sometimes think about that and I think if you then have additional information that actually shows that, that as well as that there are all these people who have got a similar virus to you, who acquired it after you. You know, I'm*

then, isn't that going to affect how I view this person, maybe not in a conscious way but…'

'So, if you then add into that additional knowledge from this, that, that is, you know, that that unprotected sex is quite possibly, you know, leading to transmission here and these actual other individuals who've recently been, erm, diagnosed with HIV, recently acquired HIV, they're having sex in the same general sort of geographical area or places, there's, there's this person that I, you know, that I started with, so I had that additional information. I'd rather not know [laughing], you know, 'cause it's, it's […] It's harder then to care for that person properly.'

Others however disagreed and were able to rationalise this on the basis of overall knowledge of the situation and of knowledge of behaviours; although patients may engage in risk behaviour they do not do so with the intention of transmitting HIV, and transmission is more likely when a person is unaware of their infection. This then raised questions as to whether it would be appropriate to inform an individual who may be implicated in transmission of this fact, given it was highly likely to be unintentional and doing so may cause undeserved self-blame.

'But there's a difference between knowingly transmitting and not knowing that you're positive and transmitting. They're two very, very different things. Even if you're taking risks and you know you're at risk, you're not, presuming, you're presuming you won't get HIV, that's why you're taking risks, because you've managed not to get HIV so far, and so you're continuing your behaviour, because it's not going be the first time you would have taken risky behaviour for the most part, it will be people who engage in risky behaviour all the time and have had HIV tests which were negative before and so they're continuing to try and get away with what they got away with before and they, are unlucky this time, they get HIV but they continue their normal behaviour because they don't know they have HIV. So it may, so I think there's a difference in someone knowingly transmitting and someone transmitting, but I think where it may be relevant to the person who then is identified as the transmitter, do we tell them, because then what impact is that going to have on them. Because the vast, I mean 99.9% of people living with HIV do not want to give anyone else HIV [...]. Most people do not anyone else to go through what they're going through, you know, even if it's a long-term condition with a normal prognosis, and life expectancy, they don't want anyone else to have to take pills or to be in their position.'

5.6. Discussion

5.6.1 Key findings

The use of phylogenetics to guide public health interventions raises complex acceptability and ethical dilemmas on a background of very complex science. Although with detailed explanation individuals with no prior knowledge appear to understand these concepts, negotiating their position on what was acceptable or not to them appears to be a challenging, and very individual process. Despite explanation, misconceptions were frequently repeated within discussions, such as the fact that phylogenetics can tell us 'who-infected-who', which may not reflect a lack of understanding but the magnitude of the possible repercussions if this was the case[322] increasing the perception of risk with phylogenetics, highlighting the difficulties in educating non-experts in this highly emotive area.

The use of phylogenetics is acceptable if participants view the benefits of use of their data to outweigh their perceived personal risk, largely of disclosure leading to further stigma, or risk of prosecution. This risk is highly dependent on personal beliefs and experiences, though several factors affecting acceptability were common throughout groups within this study. Crucial factors influencing acceptability include maintaining confidentiality through strict security measures and limitations on access to data, improved understanding and protections from use for criminalisation cases, though the weight of these protective factors varied between individuals. Nevertheless, it appears that despite safeguards, a small number of individuals will remain highly opposed to the use of phylogenetics in any way that may allow identification of an individual, and the consideration of informed consent, or an 'opt-out' option may be necessary to avoid disengaging such individuals from testing or care if the use of potentially identifiable data were considered.

5.6.2 Strengths and limitations

This study is the first to explore patient and healthcare workers views on the use of phylogenetics to guide public health interventions in HIV in the UK, and provides

valuable insight into how such interventions may be acceptably introduced in the UK. By combining data from both focus groups and interviews with different demographic groups, we were able to identify a wide range of views and explore individuals' views and concerns in depth. Although it is possible that within focus groups social desirability bias could have prompted participants to provide more accepting views of interventions that may provide a public health benefit, the addition of one-on-one interviews allowed reflective individual responses, though issues raised were similar between interview and focus group participants. It is important to note however that results of this study may not be generalisable to the entire population; participants were recruited largely from Brighton, a city with a well-informed HIV positive population, with excellent retention to care and engagement with treatment. Those recruited through community groups similarly were engaging with HIV services, and all participants were also willing to give up their time to participate in research. These participants are therefore are likely to represent highly motivated individuals who may be more willing to contribute their data to initiatives designed to prevent HIV transmission, and therefore our results may show these interventions to be more acceptable than they may be to the general population.

The individuals that would be most affected by the introduction of phylogenetically guided interventions are likely to be those newly diagnosed, not engaging in care, and possibly people who inject drugs, amongst whom recent outbreaks have been reported. Despite recruiting purposively from the two main risk groups for HIV in the UK (MSM and black African people), and including a small number of individuals who were diagnosed more recently, our results may not provide an accurate representation of people's responses to these interventions if used in real-life owing to small numbers that represent the groups most likely to be affected. Although ideally, interviews would have largely been conducted amongst these groups, due to small numbers of individuals fitting these demographics within our cohort and the participant engagement and motivation required to take part, this was not possible. Previous research exploring the use of molecular epidemiology within a research setting recruited its 21 'patient participants' through research networks, noting many were concomitantly enrolled within prevention studies, reflecting the difficulties recruiting crucial, hard to reach groups for this type of study [207].

This study took place during the 'Cambridge Analytica scandal', where the political consulting firm harvested personal data of millions of Facebook users without consent and subsequently used it for political advertising, and just prior to the introduction of

GDPR. Although phylogenetics was not directly related to Cambridge Analytica or mentioned in public reports addressing GDPR, there was a heightened awareness of 'big data', and privacy and autonomy issues relating to the use of personal data at this time. This was apparent throughout news and social media channels and appeared to create a national sense of mistrust relating to sharing of data and may have negatively influenced how participants viewed the use of their data for phylogenetics. In addition, the study took place following a high-profile HIV criminalisation case which occurred locally in Brighton. Although the term phylogenetics was not directly used in the press relating to this case, (one BBC report referred to a similarity in viruses) this may have influenced participants' responses regarding criminalisation of HIV, given this case involved apparent malicious intentional transmission to several victims.

5.6.3 Implications for current practice

Anonymous, retrospective analyses, as currently performed using resistance test data collected routinely as a by-product of clinical care were almost unanimously acceptable within this study, in light of potential public health or research benefits and lack of risk to the individual given their data is irreversibly anonymised. There were, however, concerns of risk though deductive disclosure from a small number of participants, if data is manipulated in a way to imply transmission from individuals with specific linked demographic data. Procedures carried out by the UK HIV Drug resistance database (UK RDB) are in place to mitigate risk of deductive disclosure from phylogenetic analyses in the UK, including assessment of protocols by the UK RDB steering committee with rejection, or amendment if such risks are anticipated. In addition, linked demographic data is aggregated to ensure a minimum of 200 sequences share the same characteristics. Despite these measures however, deductive disclosure may still be possible and reconsideration of this possibility is required prior to publication. Although anonymised analyses were found to be acceptable, this study raises the question of whether patients should be provided with more information at diagnosis regarding the use of their anonymised data for phylogenetics. Though some participants stated they would like to have been informed, it is not entirely clear how important this might be; almost all participants were unconcerned that their data may have already been used this way without their knowledge, and those who stated they would like to know only did so when directly asked, with a only a very small number expressing they felt informed consent

was required. Though patient representatives are involved with the UK RDB their numbers are limited, and the question of the extent of information provision required to patients may be addressed with the involvement of larger dedicated PPI groups. Nevertheless, information should be available to patients relating to the use of phylogenetics, if they wish to access it. As highlighted by this study, this is a difficult subject to clearly explain and understand, though the use of graphics and careful explanation were helpful. Although explanation by a healthcare worker in a clinical setting is possible, it is likely many professionals would struggle to concisely explain these concepts without training and dedicated time allocated. The development of an educational tool to assist with the explanation of phylogenetics is therefore crucial.

5.6.4 Implications for the design of future interventions

The protective factors identified within this study provide crucial insight into how these interventions must be designed in order to mitigate risk to individuals, and retain engagement with testing and care. One of the most important components is the design of a secure system in which only identifiable data may be accessed by the direct healthcare team. Even when using anonymised data, if phylogenetic surveillance is being performed on a large scale, measures to ensure full protection of data and strict limitations on access would be required to maintain public trust in the system, particularly in the context of a legacy of NHS data breaches and mistrust. In addition, analyses should not implicate transmission between pairs of sequences, to reduce risk of individual blame both in the eyes of the healthcare worker accessing the data, and in the 'never event' of a breach of security.

One major factor making any use of phylogenetics unacceptable was the potential of what was considered to be 'misuse' of data for criminalisation of HIV transmission. Many participants understood, after explanation, both its limitations, and value of phylogenetics in this context in terms of correctly disproving transmission, and agreed it was necessary in such cases, however use of data *collected for public health intentions* potentially being used to wrongly incriminate individuals was highly unacceptable, and may potentially lead to a reduced testing uptake[353]. Healthcare data may be obtained without consent under article 9 of the GDPR and therefore to truly remove this risk, data must be irreversibly de-anonymised, unless special governmental protections could be applied. If

change to legislation, cannot be granted, physical measures must be taken to reduce the risk of data being used to prosecute. This may be through the automated deletion of identifiers after a given period of time, allowing for investigation but no lasting risk of data misuse, or though using only fully anonymised data to guide broader interventions or to target public health messages. These strategies may reduce the utility of phylogenetics to prioritise or facilitate re-engagement into care, which may occur sometime after diagnosis, however in the UK disengagement rates are low, and the value of phylogenetics over current measures may not be great enough to warrant the storage of phylogenetically linked, potentially identifiable data.

Lastly, for acceptable use of phylogenetics in guiding public health interventions where any identifiable data is used, informed consent should be considered, or, more practically, an opt-out option. When bringing forward interventions for public health purposes including outbreak investigation, it may be argued that it is justifiable to use surveillance data without explicit consent [299, 354] given the clear public health gains. Many individuals whose data might be collected for such surveillance however will not be involved in an outbreak situation and not providing patients with autonomy of how their data is used risks damaging their trust in the healthcare system. In addition, there were a small number of individuals who were highly opposed to any use of identifiable data used within phylogenetic analyses, and patients who feel strongly that their data should not be shared outside of the HIV department may have a similar attitude. By not receiving informed consent or offering an opt-out option we risk disengaging these individuals from testing or HIV treatment, relegating the potential benefits of such a system. Consent however may be less relevant if data is collected directly for the investigation of a concerning outbreak, as suggested by most participants feeling consent is not required if the need for it to prevent spread is urgent, however this contradicted what many stated prior to the suggestion of 'urgency'. This suggests that issues revolve more around the storage of phylogenetically linked, potentially identifiable data and potential 'misuse', rather than the use of phylogenetics itself.

The processes of consent however will require development, in particular how and to what extend the complex science needs to be explained. Most individuals state they would consent to the use of their data for phylogenetics as long as appropriate safeguards were applied and explained, a key factor being what information is provided and how.

Although honesty and transparency were important, those who discussed the consent process highlighted the need to explain the purpose and potential benefits of this use of data, and stress exactly what protections and limitations for access are applied. In addition, during consent processes, it may be beneficial to explain exactly how the use of phylogenetic data may alter procedures adopted during clinical care; as many concerns raised during discussion of phylogenetically led interventions centred on the interventions themselves, which are already standard practice, rather than the role of phylogenetics in guiding them.

5.6.5 Implications for future research

The use of people's personal data is acceptable if there are significant public health benefits associated with its use. Although it is possible that phylogenetically guided interventions may have led to earlier control of the largest outbreak of HIV amongst PWID in Glasgow if routine real-time monitoring had been in place at the time[17], due to the retrospective nature of phylogenetic analyses in the UK, with a significant delay from processing of a resistance test to its availability within the UK RDB for use for phylogenetics, we are not aware of the current transmission dynamics within the UK, or to what extent 'outbreaks' that might be targeted for intervention may be occurring. It is therefore pertinent that real-time anonymous phylogenetic analysis should be piloted prior to consideration of using identifiable data. This would determine both whether there are potential targets identifiable for intervention, and whether there may be suggestion of additional benefit if the data were to be made identifiable and interventions more directly targeted. In addition, further evidence is required as to whether interventions targeted at phylogenetic clusters may have benefit, including cost-benefit which will directly influence feasibility, which we will gain with the on-going interventions being carried out in the USA by the CDC[190] (Section 1.7.3.2).

Further research is required to determine acceptability within less well-researched groups in other geographical locations, specifically within groups who inject drugs, disengage from care, and the very recently diagnosed. We identified differences in attitudes between demographic groups within this small study, and larger studies, with input from HIV community organisations and bodies such as the National AIDS Trust are required to mitigate the risk of carrying forward strategies under false assumptions of acceptability.

5.7 Conclusions

Acceptability of the use of phylogenetic data is based on risk a risk-benefit balance, which is highly individualised and differed between demographic groups. The use of sequences for anonymised analyses, as currently performed is acceptable, given results are sensitively presented, with no risk of deductive disclosure.

When designing interventions that may involve identifiable data however, phylogenetics is only acceptable if there is a significant anticipated public health benefit and certain safeguards are in place, including robust security measures, limiting access of identifiable data to the direct care team and protection from use for prosecution. Although consent may not strictly be required for outbreak investigations, this should be considered due to a small number of individuals who find this use of data unacceptable, and may be discouraged from testing or engaging in care if unable to opt out.

Chapter 6. Discussion

6.1. Introduction

During the past few years, the use of phylogenetics in HIV has advanced, and expanded as a tool to guide public health intervention throughout the USA. Being a novel means for guiding public health interventions, little evidence for its use exists, and guidelines for its effective and ethically correct use have yet to be formally developed. In the UK, the use of phylogenetics for the purposes of guiding public health interventions has only been used supportively and in a retrospective context, and its use as a primary public health tool for real-time investigations has not been widely explored. Despite a reduction in incidence of HIV in the UK over recent years, the large outbreak amongst people who inject drugs in Glasgow first identified in 2015 and other smaller outbreaks suggests the use of phylogenetics may be a useful adjunct to combination prevention strategies, particularly given the complexities of targeting more hard to reach individuals, which are likely to become the predominant sources and recipients of new infections if the downward trends in incidence continue.

This PhD therefore was designed to explore the use of phylogenetics to guide public health interventions in a manner that is acceptable within a UK context, in particular it aimed to determine;

1. How could phylogenetics be implemented in a way that is acceptable to patients
2. Whether real-time phylogenetically guided interventions be piloted on a local scale, providing evidence for effectiveness and real-life acceptability.

6.2. Key findings and unique contributions

6.2.1 Feasibility of local piloting of phylogenetically led interventions

This is the first body of work to explore the use of phylogenetics to guide real-time public health interventions in a UK context. We used phylogenetic techniques to retrospectively

analyse transmission within the Brighton HIV positive cohort, to determine whether transmissions occur locally, suggesting local piloting of real time phylogenetically guided interventions might be feasible. Within this study we designed a novel method for identifying transmission sources based on both phylogenetic and detailed clinical data, using a decision tree approach with factors known to be associated with transmission. This study demonstrated that a large proportion of transmissions are likely to occur locally but also suggested a significant contribution from non-local sources. Although identification of transmission sources does not necessarily imply they may be linked to clusters that warrant targeted public health interventions, this suggests transmission is not confined to geographical boundaries or sites of HIV care. This implies that piloting phylogenetically led interventions on a local scale would likely exclude potentially important transmission clusters meaning maximal prevention may not be possible, and national real-time surveillance is likely to be a more effective strategy. This finding directly impacts on the potential acceptability and ethical issues of using phylogenetics for public health purposes; rather than small, confined investigations, larger scale data usage with potentially huge databases may be necessary.

6.2.2 Acceptability of phylogenetically led interventions

Prior to this PhD no published research has explored the views of stakeholders on the use of real-time phylogenetics to guide public health interventions, and no research has explored their views of phylogenetics of HIV in the UK in any context.

The biggest contribution of this thesis is in providing information on how we can design acceptable interventions in the age of phylogenetics, with the knowledge that use of national datasets may be necessary for maximal impact. Acceptability is a fine balance for patients, largely centred upon confidentiality and criminalisation, balanced by potential benefits. Although the way phylogenetics is used at present in the UK appears to be acceptable owing to measures taken to anonymise data and reduce risk of deductive disclosure, many feel that more transparency is required when handling patient data in this manner. For the use of phylogeny guided public health interventions to be acceptable, evidence to support likelihood of public health benefit must be sought, robust security measures must be developed, and informed consent needs to be considered. Real challenges lie around patient and public understanding of phylogenetics and of concepts such as deductive disclosure, issues around criminalisation, and of standardly used

interventions such as contact tracing, which directly impact on acceptability where phylogenetics is used in any context.

6.3. Strengths and limitations of this thesis

6.3.1 Methodological Approach

This thesis combines phylogenetic and qualitative methods to address the questions asked, which is a rare combination owing to the vastly different skillsets required for each. The use of mixed methods within this thesis added value in two ways. Using phylogenetic techniques within chapter 4 enhanced my knowledge of the intricacies of phylogenetics, including data availability and acquisition including safeguards, and possibilities and limitations of phylogenetic methods, which allowed me to develop realistic scenarios for the later qualitative study. This deeper understanding also allowed me to explain key concepts and answer questions during interviews and group discussions. The use of phylogenetics also directly fed into the design of this study by informing how prevention strategies may be implemented; particularly the likely need to expand investigations nationally to gain maximal public health benefit. Using large amounts of data on a national, or even global scale as opposed to piloting interventions locally raises important ethical and acceptability issues relating to storage of data and data sharing that may have been overlooked if focussing scenarios locally, as proposed with the initial suggestion of a local pilot.

6.3.2 Similar research

One major strength of this thesis is the fact that little research has been conducted in this area, particularly on acceptability of phylogenetic methods within this context, making its findings particularly important. This lack of previous similar research however was also limiting; implementing phylogenetics for public health use within a UK setting would require consideration of a large number of factors that could not be covered within this body of work. This includes refining pipelines for sequencing and collation of data in near real time in the UK; exploring up to date trends in transmission using sequences from more recently infected individuals; performing cost-benefit analyses; and exploring

potential legal protections for this type of surveillance data. Without exploration of these aspects, the use of phylogenetics to guide public health interventions cannot move forward.

A wealth of research however has emerged during this PhD relating to phylogenetically guided public health interventions in the US; ethical aspects have been explored to a degree[205, 317] and methods to identify clusters where public health interventions may prevent further transmission have been identified[192]. These strategies have now been implemented throughout the USA, and though outcomes have only been touched upon within conference presentations[190], more evidence regarding its use is imminent. Unfortunately, these ethical reflections and reports were published following the design and fieldwork undertaken for the studies within this thesis; had these arisen earlier, they may have impacted directly upon them. For example, rather than assessing transmission sources within pairs phylogenetically, an alternative approach would have been to adopt the methods used in the USA to identify 'priority' clusters, if these are present. Although this may not be necessarily the best strategy to prevent transmission, given the large-scale exploration of its use, if it proves to provide prevention benefit it is likely to be adopted elsewhere and if applied within this thesis may have provided an estimation of how many similar groups in a UK population might warrant intervention. Likewise, the idea of 'priority clusters' may have been introduced into interviews and focus groups as this now appears to be a more likely direction in which the use of phylogenetics for public health interventions may be taken.

It is however important to note that interventions being undertaken in the USA based on 'priority' clusters are predicted on a particular pattern of transmission network, for which there is minimal evidence, and are likely to be very different in different risk populations. Therefore had this work been published before commencement of this PhD, rather than altering the studies as described above, a more sensible approach may have been to initially experiment with this, and other strategies with simulated phylogenetic data, such as DSPS-HIV (Discrete Spatial Phylo Simulator, modified to simulate realistic HIV epidemics)[355], customised to reflect both the UK and US epidemics. This would allow both a relatively accurate and direct comparison, to determine how relevant any future outcomes based on 'priority' clusters from the USA may be for the UK. In addition, it would provide more insight into the unsampled individuals linked to these clusters, and

therefore more evidence as to whether investigation of ‘priority’ clusters may have significant public health benefit at all.

Most subsequently published reviews of the ethical aspects of the use of phylogenetics have been based on specialist opinion, however one study provided information on the views of stakeholders, including patients, of the use of molecular epidemiology for research. Although findings were broadly similar to those identified within our qualitative study, these findings may have fed into the design of scenarios and influenced subsequent questions asked of participants had they been published earlier on.

6.3.3 Changing epidemiology

During the development of this thesis, the UK HIV epidemic has changed significantly. Combination prevention efforts have resulted in a reduction in incidence, even in the historically hard to reach group of heterosexual white men, and the proportion of undiagnosed infections has reduced. This could suggest the need for phylogenetic enhancement of prevention interventions may be less relevant than we initially thought, however the large outbreak that was concurrently reported in Glasgow suggests otherwise. It may therefore be that its utility is in the early identification and investigation of uncommon, yet potentially catastrophic outbreaks than for investigation of smaller scale investigations, though this will be clarified by surveillance of on-going trends in incidence, and reports of further outbreaks. Had these epidemiological trends been apparent earlier, this thesis may have been focussed more on the use of phylogenetics for outbreak management, particularly in chapters 2 and 3, and chapter 4 again could have focussed on rapidly growing, or large ‘priority’ clusters.

6.4. Implications for clinical practice

One immediate implication that has arisen from this thesis is the need for patient and public education relating to the use of patient data for phylogenetics. When developing scenarios and explanatory slides for interviews and discussions, there was an absence of available ‘patient friendly’ information on which to base these, meaning explanations were provided in an iterative manner, based upon misunderstandings and misconceptions

arising in previous discussions. Participants within our qualitative study expressed a desire to understand what their data was being used for, though acknowledged phylogenetic concepts were difficult to grasp. It is therefore imperative that appropriate tools are developed and made available to both patients and any other members of the public that are interested. The difficulties in describing phylogenetics within this study led to the development of a short educational YouTube animated video which is a tool that could potentially be distributed to clinics, or through organisations such as AIDSMap, or BHIVA, though resources for clinical staff must similarly be developed to ensure accurate and clear information is provided where necessary face to face.

Another significant implication from this work is that we cannot implement real-time phylogenetically guided interventions in the UK without further research. Though this may be acceptable if used in a major outbreak, such as has occurred in Glasgow, if used within the general population without a significant expected public health benefit, exploration of legal protections, informed consent and development of highly robust security measures there may be a risk of disengagement and mistrust, defeating the point of the service.

6.5. Implications for policy

Given the associations of phylogenetics with criminalisation, a major area influencing the acceptability of its use, it is critical that patients are given clear information relating to the criminalisation of HIV transmission, a topic of which understanding appears to be variable, and of which misunderstanding can heighten anxiety. BHIVA's 2013 position statement *HIV Transmission, the Law and the Work of the Clinical Team* recommends all patients diagnosed with HIV are informed of the laws surrounding reckless transmission of HIV, however this recommendation is absent from its safer sex, sexual and reproductive health, and routine investigations and monitoring guidance, meaning this may not be prioritised in clinical practice. Although discussion regarding criminalisation may seem alarming in the early stages of diagnosis, full understanding of the caveats is likely to reassure in the context of extensive media coverage in the event of cases going to trial, and being fully informed allows individuals to protect themselves from potential prosecution. It therefore would be sensible for this to become one of the auditable

outcomes within one of these major guidelines. Furthermore, in anticipation of potential phylogenetically led interventions, policy makers, in collaboration with legal experts and clinicians need to explore possible protections for phylogenetic network data used for surveillance purposes against judicial use.

Practically, one further major policy change if moving forward would be regarding the purpose of the UKRBD, which is currently a research database. Given the likely need for national background sequences, the use of almost all of the sequences within the repository may be necessary to implement successful interventions. The UKRBD is currently re-applying for ethical approval to maintain this research status, though discussions within the database committee, and with public health bodies should explore its use for surveillance purposes going ahead.

6.6. Future research and directions

Acceptability of an intervention is a key factor in its success, directly impacting on patient engagement and trust, however acceptability of phylogenetically guided interventions varies between individuals, is highly dependent on a variety of practical and personal factors, and the feasibility of these interventions does not depend on acceptability alone.

6.6.1 Prevention benefits of phylogenetically guided interventions

One major area requiring research is in the efficacy of phylogenetically guided interventions in achieving HIV prevention benefits. Analyses in the USA have identified ‘priority clusters’ with high transmission rates, though whether or not interventions have provided public health benefit is unclear, largely owing to a lack of reported outcomes due to the recency of their implementation. An up to date analysis of UK data may reveal whether such ‘priority clusters’ exist, though given the major differences in the UK epidemic when compared to the US, with high rates of testing, engagement and suppressive treatment in the UK, any growing cluster should be considered a ‘priority’, and less restrictive cut-offs as to what constitutes a priority cluster may be adopted. Given

the recent decline in incidence of HIV in the UK however, a more appropriate strategy to explore its potential in terms of identifying targets for intervention may be to perform real-time anonymised analyses to identify growing clusters, as used in British Colombia, as retrospective analyses may no longer be reflective of current transmission patterns. The utilisation of HIV-TRACE would make this computationally straightforward, requiring only development of pipelines to significantly expedite the receipt of sequences for analysis following blood being taken for resistance testing. By determining the frequency of potential targets amenable to intervention, and with future reports relating to outcomes from these interventions from the USA, we may be able to estimate the prevention effect of implementing them in the UK, which may provide insight into the cost effectiveness of phylogeny guided interventions, as well as technical and temporal challenges that may occur. Real-time analyses using anonymised data alone may even be adequate for guiding public health intervention in the future, dependent on findings; if a large growing cluster was found in a particular region, locally implemented general prevention interventions may be as effective as those targeting specific individuals, though this still risks deductive disclosure and may not be helpful in smaller 'outbreaks' in high prevalence areas such as London.

6.6.2 Patient acceptability

Concurrently, further research is required to assess acceptability. Although we identified factors which influence acceptability, this was within a specific geographical region, and views may not reflect those of others in the UK. We also presented a wide array of possibilities and scenarios given the limited practical applications at that time. With on-going practical experience and evolution of interventions in real life, these scenarios could be refined to reflect a likely model of phylogenetically guided intervention, designed with the protective factors found within our study, to gain clarity as to how acceptable this is amongst individuals from other geographical locations, and amongst those more likely to be affected, for example the very recently diagnosed.

Although expansion of our qualitative work is necessary, there may be practical difficulties recruiting individuals very shortly after diagnosis, which is the ideal time to conduct such interviews to reduce recall bias and provide a more accurate insight into the concerns of patients with this type of intervention. It is therefore pertinent that where

phylogenetics are being used in this manner (for example in the USA), nested studies to determine acceptability are conducted, and any ethical or engagement issues are accurately reported.

6.7. Closing statement

The use of phylogenetics to guide public health interventions holds promise in areas with high rates of disengagement, poor compliance or access to treatment and high rates of undiagnosed infection, and may be acceptable to patients if it delivers the prevention benefits expected in this context. Its role in HIV prevention in the UK however currently is less clear. Though Glasgow provided an opportunity to introduce this as an outbreak management tool, evidence of slowing of the epidemic suggests if it was implemented nationally, the harms may outweigh the benefits if not carefully and sensitively planned. That said, if incidence continues to fall, more intensely targeted interventions will become necessary if our aim is to end the epidemic, and transmissions will become 'critical incidents' as opposed to the accepted norm, potentially warranting phylogenetic assistance. For this reason, I believe phylogenetics as a public health tool is a key area for development in the UK, and anticipatory efforts must be made to address the complex legal, privacy, and practical challenges posed by its implementation, which if used thoughtfully in the correct context may be the missing link in our goal to reach zero.

References

1. Gottlieb MS, *et al.* Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. *N Engl J Med* **305** 1425-1431. (1981)

2. Centres for Disease Control. A Cluster of Kaposi's Sarcoma and Pneumocystis carinii Pneumonia among Homosexual Male Residents of Los Angeles and range Counties, California. *MMWR* **31** 305-7 (1982)

3. Epidemiologic aspects of the current outbreak of Kaposi's sarcoma and opportunistic infections. *N Engl J Med* **306** 248-252. (1982)

4. Global HIV & AIDS statistics — 2019 fact sheet. Available from: https://www.unaids.org/en/resources/fact-sheet (Accessed December 2019)

5. Global AIDS update 2019 — Communities at the centre: UNAIDS; 2019. Available from: http://rstesa.unaids.org/publications/global-publications/item/209-global-aids-update-2019-communities-at-the-centre (Accessed December 2019)

6. Political Declaration on HIV and AIDS: On the Fast Track to Accelerating the Fight against HIV and to Ending the AIDS Epidemic by 2030: UNAIDS; 2019. Available from: http://www.hlm2016aids.unaids.org/wp-content/uploads/2016/06/2016-political-declaration-HIV-AIDS_en.pdf. (Accessed August 2019)

7. UNAIDS. 90-90-90 An ambitious treatment target to help end the AIDS epidemic. 2014. Available from: http://www.unaids.org/sites/default/files/media_asset/90-90-90_en_0.pdf (Accessed August 2019)

8. Nash S, *et al*. Progress towards ending the HIV epidemic in the United Kingdom: 2018 report. November 2018, Public Health England, London .

9. Ragonnet-Cronin M, *et al.* Non-disclosed men who have sex with men in UK HIV transmission networks: phylogenetic analysis of surveillance data. *Lancet HIV* **5** e309-e316. (2018)

10. UNAIDS. Combination HIV Prevention: Tailoring and Coordinating Biomedical, Behavioural and Structural Strategies to Reduce New HIV Infections. (2019)

11. Ippolito G, Puro V, De Carli G. The risk of occupational human immunodeficiency virus infection in health care workers. Italian Multicenter

Study. The Italian Study Group on Occupational Risk of HIV infection. *Arch Intern Med* **153** 1451-1458. (1993)

12. Richman KM, Rickman LS. The potential for transmission of human immunodeficiency virus through human bites. *J Acquir Immune Defic Syndr* **6** 402-406. (1993)

13. Abdala N, Stephens PC, Griffith BP, Heimer R. Survival of HIV-1 in syringes. *J Acquir Immune Defic Syndr Hum Retrovirol* **20** 73-80. (1999)

14. Cresswell F, *et al.* UK guideline for the use of HIV Post-Exposure Prophylaxis Following Sexual Exposure, 2015. *Int J STD AIDS* **27** 713-738. (2016)

15. Peters PJ, *et al.* HIV Infection Linked to Injection Use of Oxymorphone in Indiana, 2014-2015. *The New England journal of medicine* **375** 229-239. (2016)

16. Paraskevis D, *et al.* Economic recession and emergence of an HIV-1 outbreak among drug injectors in Athens metropolitan area: A longitudinal study. *PLoS ONE* **8**. (2013)

17. Ragonnet-Cronin M, *et al.* Recent and Rapid Transmission of HIV among People who Inject Drugs in Scotland Revealed through Phylogenetic Analysis. *J Infect Dis*. (2018)

18. Newell ML. Mechanisms and timing of mother-to-child transmission of HIV-1. *Aids* **12** 831-837. (1998)

19. Peters H, *et al.* UK Mother-to-Child HIV Transmission Rates Continue to Decline: 2012-2014. *Clin Infect Dis* **64** 527-528. (2017)

20. 2015 progress report on the global plan towards the elimination of new HIV infections among children and keeping their mothers alive. Joint United Nations Programme on HIV/AIDS (UNAIDS); 2015.

21. Start Free. Stay Free. AIDS Free. UNAIDS. Available from: https://free.unaids.org/ (Accessed September 2019)

22. Coates TJ, Richter L, Caceres C. Behavioural strategies to reduce HIV transmission: how to make them work better. *Lancet* **372** 669-684. (2008)

23. Fox J, Fidler S. Sexual transmission of HIV-1. *Antiviral Res* **85** 276-285. (2010)

24. Giannou FK, *et al.* Condom effectiveness in reducing heterosexual HIV transmission: a systematic review and meta-analysis of studies on HIV serodiscordant couples. *Expert Rev Pharmacoecon Outcomes Res* **16** 489-499. (2016)

25. Johnson WD, O'Leary A, Flores SA. Per-partner condom effectiveness against HIV for men who have sex with men. *Aids* **32** 1499-1505. (2018)

26. Pakianathan M, *et al.* Chemsex and new HIV diagnosis in gay, bisexual and other men who have sex with men attending sexual health clinics. *HIV Med*. (2018)

27. Quinn TC, *et al.* Viral load and heterosexual transmission of human immunodeficiency virus type 1. Rakai Project Study Group. *N Engl J Med* **342** 921-929. (2000)

28. Castilla J, *et al.* Effectiveness of highly active antiretroviral therapy in reducing heterosexual transmission of HIV. *J Acquir Immune Defic Syndr* **40** 96-101. (2005)

29. Cohen MS, *et al.* Prevention of HIV-1 infection with early antiretroviral therapy. *N Engl J Med* **365** 493-505. (2011)

30. Cohen MS, *et al.* Antiretroviral Therapy for the Prevention of HIV-1 Transmission. *N Engl J Med* **375** 830-839. (2016)

31. Rodger AJ, *et al.* Sexual Activity Without Condoms and Risk of HIV Transmission in Serodifferent Couples When the HIV-Positive Partner Is Using Suppressive Antiretroviral Therapy. *Jama* **316** 171-181. (2016)

32. Rodger AJ, *et al.* Risk of HIV transmission through condomless sex in serodifferent gay couples with the HIV-positive partner taking suppressive antiretroviral therapy (PARTNER): final results of a multicentre, prospective, observational study. *Lancet*. (2019)

33. Brown AE, Gill ON, Delpech VC. HIV treatment as prevention among men who have sex with men in the UK: is transmission controlled by universal access to HIV treatment and care? *HIV Med* **14** 563-570. (2013)

34. Johnson LF, Lewis DA. The effect of genital tract infections on HIV-1 shedding in the genital tract: a systematic review and meta-analysis. *Sex Transm Dis* **35** 946-959. (2008)

35. Ghys PD, *et al.* The associations between cervicovaginal HIV shedding, sexually transmitted diseases and immunosuppression in female sex workers in Abidjan, Cote d'Ivoire. *Aids* **11** F85-93. (1997)

36. Mole L, Ripich S, Margolis D, Holodniy M. The impact of active herpes simplex virus infection on human immunodeficiency virus load. *J Infect Dis* **176** 766-770. (1997)

37. Levine WC, *et al.* Increase in endocervical CD4 lymphocytes among women with nonulcerative sexually transmitted diseases. *J Infect Dis* **177** 167-174. (1998)

38. Taha TE, *et al.* Bacterial vaginosis and disturbances of vaginal flora: association with increased acquisition of HIV. *Aids* **12** 1699-1706. (1998)

39. UK National Guidelines for HIV Testing 2008. Available from: http://www.bhiva.org/documents/guidelines/testing/glineshivtest08.pdf (Accessed February 2019)

40. Auvert B, *et al.* Randomized, controlled intervention trial of male circumcision for reduction of HIV infection risk: the ANRS 1265 Trial. *PLoS Med* **2** e298. (2005)

41. Gray RH, *et al.* Male circumcision for HIV prevention in men in Rakai, Uganda: a randomised trial. *Lancet* **369** 657-666. (2007)

42. Millett GA, Flores SA, Marks G, Reed JB, Herbst JH. Circumcision status and risk of HIV and sexually transmitted infections among men who have sex with men: a meta-analysis. *Jama* **300** 1674-1684. (2008)

43. Jin F, *et al.* Per-contact probability of HIV transmission in homosexual men in Sydney in the era of HAART. *Aids* **24** 907-913. (2010)

44. McCoombe SG, Short RV. Potential HIV-1 target cells in the human penis. *Aids* **20** 1491-1495. (2006)

45. Donoval BA, *et al.* HIV-1 target cells in foreskins of African men with varying histories of sexually transmitted infections. *Am J Clin Pathol* **125** 386-391. (2006)

46. Phillips AN, *et al.* Potential impact on HIV incidence of higher HIV testing rates and earlier antiretroviral therapy initiation in MSM. *Aids* **29** 1855-1862. (2015)

47. Kennedy CE, Medley AM, Sweat MD, O'Reilly KR. Behavioural interventions for HIV positive prevention in developing countries: a systematic review and meta-analysis. *Bull World Health Organ* **88** 615-623. (2010)

48. LaCroix JM, Pellowski JA, Lennon CA, Johnson BT. Behavioural interventions to reduce sexual risk for HIV in heterosexual couples: a meta-analysis. *Sex Transm Infect* **89** 620-627. (2013)

49. Ross DA. Behavioural interventions to reduce HIV risk: what works? *Aids* **24 Suppl 4** S4-14. (2010)

50. Gazzard BG, *et al.* British HIV Association Guidelines for the treatment of HIV-1-infected adults with antiretroviral therapy 2008. *HIV Med* **9** 563-608. (2008)

51. Lundgren JD, *et al.* Initiation of Antiretroviral Therapy in Early Asymptomatic HIV Infection. *N Engl J Med* **373** 795-807. (2015)

52. Otten RA, *et al.* Efficacy of postexposure prophylaxis after intravaginal exposure of pig-tailed macaques to a human-derived retrovirus (human immunodeficiency virus type 2). *J Virol* **74** 9771-9775. (2000)

53. Cardo DM, *et al.* A case-control study of HIV seroconversion in health care workers after percutaneous exposure. Centers for Disease Control and Prevention Needlestick Surveillance Group. *N Engl J Med* **337** 1485-1490. (1997)

54. Wade NA, *et al.* Abbreviated regimens of zidovudine prophylaxis and perinatal transmission of the human immunodeficiency virus. *N Engl J Med* **339** 1409-1414. (1998)

55. McCormack S, *et al.* Pre-exposure prophylaxis to prevent the acquisition of HIV-1 infection (PROUD): effectiveness results from the pilot phase of a pragmatic open-label randomised trial. *Lancet* **387** 53-60. (2016)

56. Molina JM, *et al.* On-Demand Preexposure Prophylaxis in Men at High Risk for HIV-1 Infection. *N Engl J Med* **373** 2237-2246. (2015)

57. PrEPster, iwantprepnow, and Public Health England.PrEP User May 2018 Online Survey-Summary Results. 2018; Available from:www.aidsmap.com/Nearly-a-quarter-of-people-who-want-PrEP-currently-cant-get-it-UK-survey-finds/page/3297439/. (Accessed August 2018)

58. Angus BB, Gary. Awosusi, Funmi. Barker, Gary. BHIVA guidelines for the routine investigation and monitoring of adult HIV-1-positive individuals 2016 (2019 interim update). BHIVA 2019. Available from: https://www.bhiva.org/monitoring-guidelines (Accessed August 2019)

59. The Manual for Sexual Health Advisors. Society of Sexual Health Advisors. 2004. Available from: http://www.ssha.info/wp-content/uploads/ha_manual_2004_complete.pdf (Accessed May 2017)

60. Rayment M, *et al.* An effective strategy to diagnose HIV infection: findings from a national audit of HIV partner notification outcomes in sexual health and infectious disease clinics in the UK. *Sex Transm Infect* **93** 94-99. (2017)

61. HIV Partner Notification: a missed opportunity? National AIDS Trust Report, May 2012. Available from: http://www.bhiva.org/documents/Publications/May-2012-HIV-Partner-Notification.pdf. (Accessed May 2017)

62. Erens BM, *et al.* National Survey of Sexual Attitudes and Lifestyles II. Reference tables and summary report. 2003. Available from: http://natsal.ac.uk/media/2083/reference_tables_and_summary_report.pdf (Accessed May 2017)

63. Mercer CH, Aicken CR, Brook MG, Estcourt CS, Cassell JA. Estimating the likely public health impact of partner notification for a clinical service: an evidence-based algorithm. *American journal of public health* **101** 2117-2123. (2011)

64. Moore ZS, McCoy S, Kuruc J, Hilton M, Leone P. Number of named partners and number of partners newly diagnosed with HIV infection identified by persons with acute versus established HIV infection. *J Acquir Immune Defic Syndr* **52** 509-513. (2009)

65. Liau A, Millett G, Marks G. Meta-analytic examination of online sex-seeking and sexual risk behavior among men who have sex with men. *Sex Transm Dis* **33** 576-584. (2006)

66. Klein H. Anonymous sex and HIV risk practices among men using the Internet specifically to find male partners for unprotected sex. *Public Health* **126** 471-481. (2012)

67. Sanford D. Back to a Future: One Man's AIDS Tale Shows How Quickly Epidemic Has Turned. *Oncologist* **2** 115-120. (1997)

68. Herek GM, Capitanio JP, Widaman KF. HIV-related stigma and knowledge in the United States: prevalence and trends, 1991-1999. *Am J Public Health* **92** 371-377. (2002)

69. Ross JD, Scott GR. The association between HIV media campaigns and number of patients coming forward for HIV antibody testing. *Genitourin Med* **69** 193-195. (1993)

70. Slavin S, Batrouney C, Murphy D. Fear appeals and treatment side-effects: an effective combination for HIV prevention? *AIDS Care* **19** 130-137. (2007)

71. Sparrowhawk A. Perceptions of HIV within the general public. Session presented at: 23rd Annual Conference of the British HIV Association; 2017; Liverpool. Available from: https://www.bhiva.org/file/hAgNkocyELzrp/AlexSparrowHawk.pdf

72. Lorenc T, *et al.* HIV testing among men who have sex with men (MSM): systematic review of qualitative evidence. *Health Educ Res* **26** 834-846. (2011)

73. Gwadz M, *et al.* Doing battle with "the monster:" how high-risk heterosexuals experience and successfully manage HIV stigma as a barrier to HIV testing. *Int J Equity Health* **17** 46. (2018)

74. Daskalopoulou M, *et al.* Non-Disclosure of HIV Status and Associations with Psychological Factors, ART Non-Adherence, and Viral Load Non-Suppression Among People Living with HIV in the UK. *AIDS Behav* **21** 184-195. (2017)

75. Shubber Z, *et al.* Patient-Reported Barriers to Adherence to Antiretroviral Therapy: A Systematic Review and Meta-Analysis. *PLoS Med* **13**. (2016)

76. Sayles JN, Wong MD, Kinsler JJ, Martins D, Cunningham WE. The association of stigma with self-reported access to medical care and antiretroviral therapy adherence in persons living with HIV/AIDS. *J Gen Intern Med* **24** 1101-1108. (2009)

77. Zwalen M, Egger M. Progression and mortality of untreated HIV-positive individuals living in resource-limited settings: Update of literature review and evidence synthesis. Geneva: UNAIDS; 2006.

78. Pantaleo G, Graziosi C, Fauci AS. The immunopathogenesis of human immunodeficiency virus infection. *N Engl J Med* **328** 327-335. (1993)

79. Revised surveillance case definition for HIV infection--United States, 2014. *MMWR Recomm Rep* **63** 1-10. (2014)

80. WHO case definitions of HIV for surveillance and revised clinical staging and immunological classification of HIV-related disease in adults and children. *WHO*. (2015)

81. Kassutto S, Rosenberg ES. Primary HIV type 1 infection. *Clin Infect Dis* **38** 1447-1453. (2004)

82. Kahn JO, Walker BD. Acute human immunodeficiency virus type 1 infection. *N Engl J Med* **339** 33-39. (1998)

83. Robb ML, *et al.* Prospective Study of Acute HIV-1 Infection in Adults in East Africa and Thailand. *N Engl J Med* **374** 2120-2130. (2016)

84. Daar ES, Pilcher CD, Hecht FM. Clinical presentation and diagnosis of primary HIV-1 infection. *Curr Opin HIV AIDS* **3** 10-15. (2008)

85. Garrett N, *et al.* The Recent Infection Testing Algorithm (RITA) in clinical practice: a survey of HIV clinicians in England and Northern Ireland. *HIV medicine* **13** 444-447. (2012)

86. Aghaizu A, *et al.* Recent infection testing algorithm (RITA) applied to new HIV diagnoses in England, Wales and Northern Ireland, 2009 to 2011. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin* **19**. (2014)

87. Carlin E, Taha Y. Using recent infection testing algorithm tests in clinical practice. *Sexually transmitted infections* **88** 304-306. (2012)

88. Fisher M, *et al.* Determinants of HIV-1 transmission in men who have sex with men: A combined clinical, epidemiological and phylogenetic approach. *AIDS* **24** 1739-1747. (2010)

89. Wawer MJ, *et al.* Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda. *Journal of Infectious Diseases* **191** 1403-1409. (2005)

90. Brenner BG, *et al.* High rates of forward transmission events after acute/early HIV-1 infection. *J Infect Dis* **195** 951-959. (2007)

91. Kirwan PD, *et al.* HIV in the UK - 2016 report. Public Health England, London. (2016)

92. Saez-Cirion A, *et al.* Post-treatment HIV-1 controllers with a long-term virological remission after the interruption of early initiated antiretroviral therapy ANRS VISCONTI Study. *PLoS Pathog* **9** e1003211. (2013)

93. Blankson JN, Persaud D, Siliciano RF. The challenge of viral reservoirs in HIV-1 infection. *Annu Rev Med* **53** 557-593. (2002)

94. Thornhill J, Fidler S, Frater J. Advancing the HIV cure agenda: the next 5 years. *Curr Opin Infect Dis* **28** 1-9. (2015)

95. Mellors JW, *et al.* Quantitation of HIV-1 RNA in plasma predicts outcome after seroconversion. *Ann Intern Med* **122** 573-579. (1995)

96. Pantaleo G, Fauci AS. Immunopathogenesis of HIV infection. *Annu Rev Microbiol* **50** 825-854. (1996)

97. Yarchoan R, *et al.* CD4 count and the risk for death in patients infected with HIV receiving antiretroviral therapy. *Ann Intern Med* **115** 184-189. (1991)

98. Lundgren JD, *et al.* Comparison of long-term prognosis of patients with AIDS treated and not treated with zidovudine. AIDS in Europe Study Group. *Jama* **271** 1088-1092. (1994)

99. Delta: a randomised double-blind controlled trial comparing combinations of zidovudine plus didanosine or zalcitabine with zidovudine alone in HIV-infected individuals. Delta Coordinating Committee. *Lancet* **348** 283-291. (1996)

100. Hammer SM, *et al.* A controlled trial of two nucleoside analogues plus indinavir in persons with human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less. AIDS Clinical Trials Group 320 Study Team. *N Engl J Med* **337** 725-733. (1997)

101. May M, *et al.* Impact of late diagnosis and treatment on life expectancy in people with HIV-1: UK Collaborative HIV Cohort (UK CHIC) Study. *Bmj* **343** d6016. (2011)

102. Nakagawa F, *et al.* Projected life expectancy of people with HIV according to timing of diagnosis. *Aids* **26** 335-343. (2012)

103. Survival of HIV-positive patients starting antiretroviral therapy between 1996 and 2013: a collaborative analysis of cohort studies. *Lancet HIV*. (2017)

104. HIV drug resistance in the UK. UK HIV Drug Resistance Database. 2019 Available from: http://www.hivrdb.org.uk/hiv-drug-resistance-uk (Accessed June 2019)

105. Popovic M, Sarngadharan MG, Read E, Gallo RC. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. *Science* **224** 497-500. (1984)

106. Levy JA, *et al.* Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS. *Science* **225** 840-842. (1984)

107. Coffin J, *et al.* What to call the AIDS virus? *Nature* **321** 10. (1986)

108. Sharp PM, Hahn BH. Origins of HIV and the AIDS Pandemic. *Cold Spring Harb Perspect Med* **1**. (2011)

109. Huet T, Cheynier R, Meyerhans A, Roelants G, Wain-Hobson S. Genetic organization of a chimpanzee lentivirus related to HIV-1. *Nature* **345** 356-359. (1990)

110. Hirsch VM, Olmsted RA, Murphey-Corb M, Purcell RH, Johnson PR. An African primate lentivirus (SIVsm) closely related to HIV-2. *Nature* **339** 389-392. (1989)

111. Faria NR, *et al.* HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. *Science (New York, NY)* **346** 56-61. (2014)

112. Worobey M, *et al.* 1970s and 'Patient 0' HIV-1 genomes illuminate early HIV/AIDS history in North America. *Nature* **539** 98-101. (2016)

113. Novitsky V, Wang R, Lagakos S, Essex M. HIV-1 Subtype C Phylodynamics in the Global Epidemic. *Viruses* **2** 33-54. (2010)

114. Neogi U, *et al.* Molecular epidemiology of HIV-1 subtypes in India: origin and evolutionary history of the predominant subtype C. *PLoS One* **7** e39819. (2012)

115. de Oliveira T, Pillay D, Gifford RJ. The HIV-1 subtype C epidemic in South America is linked to the United Kingdom. *PLoS One* **5** e9311. (2010)

116. Kariuki SM, Selhorst P, Ariën KK, Dorfman JR. The HIV-1 transmission bottleneck. *Retrovirology* **14** 1-19. (2017)

117. Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science* **271** 1582-1586. (1996)

118. Arts EJ, Hazuda DJ. HIV-1 Antiretroviral Drug Therapy. *Cold Spring Harb Perspect Med* **2**. (2012)

119. Charpentier C, Nora T, Tenaillon O, Clavel F, Hance AJ. Extensive recombination among human immunodeficiency virus type 1 quasispecies makes an important contribution to viral diversity in individual patients. *J Virol* **80** 2472-2482. (2006)

120. Nowak MA. What is a quasispecies? *Trends Ecol Evol* **7** 118-121. (1992)

121. Robertson DL, *et al.* HIV-1 nomenclature proposal. *Science* **288** 55-56. (2000)

122. Gurtler LG, *et al.* A new subtype of human immunodeficiency virus type 1 (MVP-5180) from Cameroon. *J Virol* **68** 1581-1585. (1994)

123. De Leys R, *et al.* Isolation and partial characterization of an unusual human immunodeficiency retrovirus from two persons of west-central African origin. *J Virol* **64** 1207-1216. (1990)

124. Plantier JC, *et al.* A new human immunodeficiency virus derived from gorillas. *Nat Med* **15** 871-872. (2009)

125. Kuiken C L, *et al.* Human retroviruses and AIDS 1999 : a compilation and analysis of nucleic acid and amino acid sequences . Los Alamos, N.M: Theoretical Biology and Biophysics Group, Los Alamos National Laboratory. 492–505. (1999)

126. Lemey P, Rambaut A, Pybus OG. HIV evolutionary dynamics within and among hosts. *AIDS Rev* **8** 125-140. (2006)

127. Lemey P, Salemi M, Vandamme A-M. *The Phylogenetic Handbook. A practical approach to phylogenetic analysis and hypothesis testing.*, 2nd edn. Cambridge University Press: New York, 2009.

128. Novitsky V, Moyo S, Lei Q, Degruttola V, Essex M. Impact of sampling density on the extent of HIV clustering. *AIDS Research and Human Retroviruses* **30** 1226-1235. (2014)

129. Eigen M, Schuster P. *The Hypercycle; A principle of Natural Self-Organization*. Springer-Verlag: Berlin, 1979.

130. Hue S, Clewley JP, Cane PA, Pillay D. HIV-1 pol gene variation is sufficient for reconstruction of transmissions in the era of antiretroviral therapy. *AIDS* **18** 719-728. (2004)

131. Mbisa JL, *et al.* Determining the origins of HIV-1 drug-resistant minority variants in people who are recently infected using phylogenetic reconstruction. *Clin Infect Dis*. (2018)

132. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25** 4876-4882. (1997)

133. Notredame C, Higgins DG, Heringa J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302** 205-217. (2000)

134. Hall T. BioEdit. Available at http://www.mbio.ncsu.edu/BioEdit/bioedit.html.

135. GeneiousAvailable from: https://www.geneious.com

136. Jukes THC, C R. Evolution of protein molecules. In: Munroe HH (ed). *Mammalian Protein Metabolism.*, vol. III. Academic Press: New York, 1969, pp 21-132.

137. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* **16** 111-120. (1980)

138. Felsenstein J. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* **17** 368-376. (1981)

139. Hasegawa M, Kishino H, Yano T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* **22** 160-174. (1985)

140. Tavaré./ S, Miura RM (eds). *Some probabilistic and statistical problems in the analysis of DNA sequences*. American Mathematical Society, 1986.

141. Shoemaker JS, Fitch WM. Evidence from nuclear sequences that invariable sites should be considered when sequence divergence is calculated. *Mol Biol Evol* **6** 270-289. (1989)

142. Yang Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J Mol Evol* **39** 306-314. (1994)

143. Sumner JG, *et al.* Is the general time-reversible model bad for molecular phylogenetics? *Syst Biol* **61** 1069-1074. (2012)

144. Posada D, Crandall KA. Selecting models of nucleotide substitution: an application to human immunodeficiency virus 1 (HIV-1). *Mol Biol Evol* **18** 897-906. (2001)

145. Hassan AS, Pybus OG, Sanders EJ, Albert J, Esbjornsson J. Defining HIV-1 transmission clusters based on sequence data. *Aids* **31** 1211-1222. (2017)

146. Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14** 817-818. (1998)

147. Zuckerkandl DJ, Pauling L. Molecular disease, evolution, and genetic heterogeneity. In: Kasha M, Pullman B (eds). *Horizons in Biochemistry*. Academic Press: New York, 1962, pp 189-225.

148. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biol* **4** e88. (2006)

149. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4** 406-425. (1987)

150. Felsenstein J. Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods Enzymol* **266** 418-427. (1996)

151. Efron B, Halloran E, Holmes S. Bootstrap confidence levels for phylogenetic trees. *Proc Natl Acad Sci U S A* **93** 7085-7090. (1996)

152. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39** 783-791. (1985)

153. Wertheim JO, *et al.* The global transmission network of HIV-1. *Journal of Infectious Diseases* **209** 304-313. (2014)

154. Ragonnet-Cronin M, *et al.* Automated analysis of phylogenetic clusters. *BMC Bioinformatics* **14** 317. (2013)

155. Prosperi MCF, *et al.* A novel methodology for large-scale phylogeny partition. *Nature Communications* **2**. (2011)

156. Kosakovsky Pond SL, Weaver S, Leigh Brown AJ, Wertheim JO. HIV-TRACE (TRAnsmission Cluster Engine): a Tool for Large Scale Molecular Epidemiology of HIV-1 and Other Rapidly Evolving Pathogens. *Mol Biol Evol* **35** 1812-1819. (2018)

157. Keele BF, *et al.* Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proceedings of the National Academy of Sciences of the United States of America* **105** 7552-7557. (2008)

158. Li H, *et al.* High multiplicity infection by HIV-1 in men who have sex with men. *PLoS Pathogens* **6** 1-17. (2010)

159. Romero-Severson EO, Bulla I, Leitner T. Phylogenetically resolving epidemiologic linkage. *Proceedings of the National Academy of Sciences of the United States of America* **113** 2690-2695. (2016)

160. Rose R, *et al.* Phylogenetic methods inconsistently predict direction of HIV transmission among heterosexual pairs in the HPTN052 cohort. *J Infect Dis*. (2018)

161. Ratmann O, *et al.* Inferring HIV-1 transmission networks and sources of epidemic spread in Africa with deep-sequence phylogenetic analysis. *Nat Commun* **10** 1411. (2019)

162. Shankarappa R, *et al.* Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J Virol* **73** 10489-10502. (1999)

163. Pond SL, *et al.* Adaptation to different human populations by HIV-1 revealed by codon-based analyses. *PLoS Comput Biol* **2** e62. (2006)

164. Poon AF, *et al.* Adaptation to human populations is revealed by within-host polymorphisms in HIV-1 and hepatitis C virus. *PLoS Pathog* **3** e45. (2007)

165. Posada D, Crandall KA, Holmes EC. Recombination in evolutionary genomics. *Annu Rev Genet* **36** 75-97. (2002)

166. Robertson DL, Hahn BH, Sharp PM. Recombination in AIDS viruses. *J Mol Evol* **40** 249-259. (1995)

167. Keele BF, *et al.* Characterization of the follicular dendritic cell reservoir of human immunodeficiency virus type 1. *J Virol* **82** 5548-5561. (2008)

168. Buzon MJ, *et al.* Deep molecular characterization of HIV-1 dynamics under suppressive HAART. *PLoS Pathog* **7** e1002314. (2011)

169. Hue S, *et al.* Phylogenetic analyses reveal HIV-1 infections between men misclassified as heterosexual transmissions. *AIDS* **28** 1967-1975. (2014)

170. Dennis AM, *et al.* Phylogenetic insights into regional HIV transmission. *AIDS* **26** 1813-1822. (2012)

171. Aldous JL, *et al.* Characterizing HIV transmission networks across the United States. *Clin Infect Dis* **55** 1135-1143. (2012)

172. Kouyos RD, *et al.* Molecular epidemiology reveals long-term changes in HIV type 1 subtype B transmission in Switzerland. *Journal of Infectious Diseases* **201** 1488-1497. (2010)

173. Brenner BG, *et al.* Transmission clustering drives the onward spread of the HIV epidemic among men who have sex with men in Quebec. *J Infect Dis* **204** 1115-1119. (2011)

174. Ou CY, *et al.* Molecular epidemiology of HIV transmission in a dental practice. *Science* **256** 1165-1171. (1992)

175. Jaffe HW, *et al.* Lack of HIV transmission in the practice of a dentist with AIDS. *Annals of Internal Medicine* **121** 855-859. (1994)

176. Hayman A, *et al.* Phylogenetic analysis of multiple heterosexual transmission events involving subtype B of HIV type 1. *AIDS Research and Human Retroviruses* **17** 689-695. (2001)

177. Allen KW, *et al.* Doncaster: the public health response to a local cluster of heterosexually acquired HIV infection. *Commun Dis Public Health* **5** 271-275. (2002)

178. Yirrell DL, *et al.* Molecular investigation into outbreak of HIV in a Scottish prison. *British Medical Journal* **314** 1446-1450. (1997)

179. Paraskevis D, *et al.* Enhanced HIV-1 surveillance using molecular epidemiology to study and monitor HIV-1 outbreaks among intravenous drug users (IDUs) in Athens and Bucharest. *Infect Genet Evol* **35** 109-121. (2015)

180. Peters PJ, *et al.* HIV Infection Linked to Injection Use of Oxymorphone in Indiana, 2014-2015. *N Engl J Med* **375** 229-239. (2016)

181. Robineau O, *et al.* Combining the Estimated Date of HIV Infection with a Phylogenetic Cluster Study to Better Understand HIV Spread: Application in a Paris Neighbourhood. *PLoS One* **10** e0135367. (2015)

182. Volz EM, Koopman JS, Ward MJ, Brown AL, Frost SD. Simple epidemiological dynamics explain phylogenetic clustering of HIV from patients with recent infection. *PLoS Comput Biol* **8** e1002552. (2012)

183. Lewis F, Hughes GJ, Rambaut A, Pozniak A, Leigh AJ. Episodic sexual transmission of HIV revealed by molecular phylodynamics. *PLoS medicine* **5**. (2008)

184. Poon AF, *et al.* The impact of clinical, demographic and risk factors on rates of HIV transmission: a population-based phylogenetic analysis in British Columbia, Canada. *J Infect Dis* **211** 926-935. (2015)

185. Chalmet K, *et al.* Epidemiological study of phylogenetic transmission clusters in a local HIV-1 epidemic reveals distinct differences between subtype B and non-B infections. *BMC Infectious Diseases* **10**. (2010)

186. Leigh Brown AJ, *et al.* Transmission network parameters estimated from HIV sequences for a nationwide epidemic. *The Journal of infectious diseases* **204** 1463-1469. (2011)

187. Smith DM, *et al.* A public health model for the molecular surveillance of HIV transmission in San Diego, California. *AIDS* **23** 225-232. (2009)

188. Little SJ, *et al.* Using HIV networks to inform real time prevention interventions. *PLoS ONE* **9**. (2014)

189. Poon FY, *et al.* Near real-time monitoring of HIV transmission hotspots from routine HIV genotyping: an implementation case study. *The lancet HIV* **3**. (2016)

190. Oster AM. Hugging phylogenetic trees: use of molecular analysis for public health intervention. Session presented at: Conference on Retroviruses and Opportunistic Infections 2019; March 4-7, 2019; Seattle, WA. Available from: http://www.croiconference.org/sessions/hugging-phylogenetic-trees-use-molecular-analysis-public-health-intervention .

191. Fauci AS, Redfield RR, Sigounas G, Weahkee MD, Giroir BP. Ending the HIV Epidemic: A Plan for the United States. *Jama*. (2019)

192. Oster AM, *et al.* Identifying Clusters of Recent and Rapid HIV Transmission Through Analysis of Molecular Surveillance Data. *J Acquir Immune Defic Syndr* **79** 543-550. (2018)

193. Public Health England. Shooting Up: Infections among people who inject drugs in the UK, 2016, An update, November 2017. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/663003/Shooting_Up_2017_report.pdf. (Accessed February 2019)

194. News: Rise in recently acquired HIV in People who inject Drugs in Dublin - Health Protection Surveillance Centre. 2015. Available from: http://www.hpsc.ie/a-z/hivstis/hivandaids/news/title-15231-en.html. (Accessed February 2019)

195. Phillips AN, *et al.* Increased HIV incidence in men who have sex with men despite high levels of ART-induced viral suppression: analysis of an extensively documented epidemic. *PLoS One* **8** e55312. (2013)

196. Ragonnet-Cronin M, *et al.* Recent and Rapid Transmission of HIV Among People Who Inject Drugs in Scotland Revealed Through Phylogenetic Analysis. *J Infect Dis* **217** 1875-1882. (2018)

197. Paraskevis D, *et al.* HIV-1 outbreak among injecting drug users in Greece, 2011: A preliminary report. *Eurosurveillance* **16**. (2011)

198. Blick G, *et al.* The probable source of both the primary multidrug-resistant (MDR) HIV-1 strain found in a patient with rapid progression to AIDS and a second recombinant MDR strain found in a chronically HIV-1-infected patient. *Journal of Infectious Diseases* **195** 1250-1260. (2007)

199. Centers for Disease Control. Investigation of a new diagnosis of multidrug-resistant, dual-tropic HIV-1 infection--New York City, 2005. *MMWR Morbidity and mortality weekly report* **55** 793. (2006)

200. Katchman E, *et al.* Successful control of a large outbreak of HIV infection associated with injection of cathinone derivatives in Tel Aviv, Israel. *Clin Microbiol Infect* **23** 336.e335-336.e338. (2017)

201. Robbins KE, *et al.* Molecular analysis in support of an investigation of a cluster of HIV-1-infected women. *AIDS Res Hum Retroviruses* **18** 1157-1161. (2002)

202. Buskin SE, *et al.* Large phylogenetically linked HIV cluster in king county, Washington, 2008 to 2014. *Topics in Antiviral Medicine* **23**. (2015)

203. Dennis A, Hue S, Sebastian J, Mobley V, Miller Wea. HIV Phylodynamics in North Carolina: Detecting Active Clusters for Intervention. Presented at: Conference on Retroviruses and Opportunistic Infections 2016; February 22-25, 2016; Boston, MA

204. Oster AM, France AM, Mermin J. Molecular Epidemiology and the Transformation of HIV Prevention. *Jama* **319** 1657-1658. (2018)

205. Mutenherwa F, Wassenaar DR, de Oliveira T. Ethical issues associated with HIV phylogenetics in HIV transmission dynamics research: A review of the literature using the Emanuel Framework. *Dev World Bioeth*. (2018)

206. Poon AFY, Dearlove BL. Quantifying the Aftermath: Recent Outbreaks Among People Who Inject Drugs and the Utility of Phylodynamics. *J Infect Dis* **217** 1854-1857. (2018)

207. Schairer C, *et al.* Perceptions of molecular epidemiology studies of HIV among stakeholders. *J Public Health Res* **6** 992. (2017)

208. Geller G, *et al.* Genomics and infectious disease: a call to identify the ethical, legal and social implications for public health and clinical practice. *Genome Med* **6** 106. (2014)

209. Fanoy E, National Institute for Public Health and the Environment tN, De Neeling A, National Institute for Public Health and the Environment tN. Molecular Typing: Use with Care. *Public Health Ethics* **5** 313-314. (2017)

210. Bernard EJ, Azad Y, Vandamme AM, Weait M, Geretti AM. HIV forensics: pitfalls and acceptable standards in the use of phylogenetic analysis as evidence in criminal investigations of HIV transmission. *HIV Med* **8** 382-387. (2007)

211. Oaten M, Stevenson RJ, Case TI. Disease avoidance as a functional basis for stigmatization. *Philos Trans R Soc Lond B Biol Sci* **366** 3433-3452. (2011)

212. Craig GM, Daftary A, Engel N, O'Driscoll S, Ioannaki A. Tuberculosis stigma as a social determinant of health: a systematic mapping review of research in low incidence countries. *Int J Infect Dis*. (2016)

213. Mak WW, *et al.* Comparative stigma of HIV/AIDS, SARS, and tuberculosis in Hong Kong. *Soc Sci Med* **63** 1912-1922. (2006)

214. TB strain typing and cluster investigation handbook: third edition, February 2014. [cited 2018 07/06/2018]Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/433055/TB_Strain_Typing_Handbook_final_070214__2_.pdf

215. Faccini M, *et al.* Tuberculosis outbreak in a primary school, Milan, Italy. *Emerg Infect Dis* **19** 485-487. (2013)

216. Bifani PJ, *et al.* Identification of a W variant outbreak of Mycobacterium tuberculosis via population-based molecular epidemiology. *Journal of the American Medical Association* **282** 2321-2327. (1999)

217. Merritt TD, *et al.* An outbreak of pulmonary tuberculosis in young Australians. *Medical Journal of Australia* **186** 240-242. (2007)

218. Lambregts-van Weezenbeek CS, *et al.* Tuberculosis contact investigation and DNA fingerprint surveillance in The Netherlands: 6 years' experience with nation-wide cluster feedback and cluster monitoring. *Int J Tuberc Lung Dis* **7** S463-470. (2003)

219. Ashworth M, *et al.* Use of PCR-based Mycobacterium tuberculosis genotyping to prioritize tuberculosis outbreak control activities. *Journal of Clinical Microbiology* **46** 856-862. (2008)

220. Clark CM, *et al.* Universal genotyping in Tuberculosis Control Program, New York City, 2001-2003. *Emerging Infectious Diseases* **12** 719-724. (2006)

221. Ruddy MC, *et al.* Outbreak of isoniazid resistant tuberculosis in north London. *Thorax* **59** 279-285. (2004)

222. Sebek M. DNA fingerprinting and contact investigation. *Int J Tuberc Lung Dis* **4** S45-48. (2000)

223. de G, van RAH, Burdo CCA, van D, Richardus JH. A Mycobacterium tuberculosis cluster demonstrating the use of genotyping in urban tuberculosis control. *BMC Infectious Diseases* **9**. (2009)

224. Malakmadze N, *et al.* Unsuspected recent transmission of tuberculosis among high-risk groups: Implications of universal tuberculosis genotyping in its detection. *Clinical Infectious Diseases* **40** 366-373. (2005)

225. Kiers A, Drost AP, van Soolingen D, Veen J. Use of DNA fingerprinting in international source case finding during a large outbreak of tuberculosis in The Netherlands. *Int J Tuberc Lung Dis* **1** 239-245. (1997)

226. Black AT, *et al.* Tracking and responding to an outbreak of tuberculosis using MIRU-VNTR genotyping and whole genome sequencing as epidemiological tools. *J Public Health (Oxf)* 1-8. (2017)

227. Ho ZJM, *et al.* Investigation of a cluster of multi-drug resistant tuberculosis in a high-rise apartment block in Singapore. *Int J Infect Dis* **67** 46-51. (2018)

228. Lalor MK, *et al.* The use of whole-genome sequencing in cluster investigation of an MDR-TB outbreak. *Eur Respir J*. (2018)

229. Walker TM, *et al.* A cluster of multidrug-resistant Mycobacterium tuberculosis among patients arriving in Europe from the Horn of Africa: a molecular epidemiological study. *Lancet Infect Dis* **18** 431-440. (2018)

230. Tuberculosis outbreak associated with a homeless shelter - Kane County, Illinois, 2007-2011. *MMWR Morb Mortal Wkly Rep* **61** 186-189. (2012)

231. Tuberculosis transmission in a homeless shelter population--New York, 2000-2003. *MMWR Morb Mortal Wkly Rep* **54** 149-152. (2005)

232. Asghar RJ, *et al.* Limited utility of name-based tuberculosis contact investigations among persons using illicit drugs: Results of an outbreak investigation. *Journal of Urban Health* **86** 776-780. (2009)

233. Bloss E, *et al.* Challenges and opportunities in a tuberculosis outbreak investigation in Southern Mississippi, 2005-2007. *Southern Medical Journal* **104** 731-735. (2011)

234. Lofy KH, *et al.* Outbreak of tuberculosis in a homeless population involving multiple sites of transmission. *International Journal of Tuberculosis and Lung Disease* **10** 683-689. (2006)

235. Investigation of a genotype cluster of tuberculosis cases -- Detroit, Michigan, 2004-2007. *MMWR: Morbidity & Mortality Weekly Report* **58** 226-230. (2009)

236. Mitruka K, *et al.* A tuberculosis outbreak fueled by cross-border travel and illicit substances: Nevada and Arizona. *Public Health Reports* **129** 78-85. (2014)

237. Miller AC, *et al.* Impact of genotyping of Mycobacterium tuberculosis on public health practice in Massachusetts. *Emerging Infectious Diseases* **8** 1285-1289. (2002)

238. Notes from the field: tuberculosis cluster associated with homelessness - duval county, Florida, 2004-2012. *MMWR: Morbidity & Mortality Weekly Report* **61** 539-541. (2012)

239. Buff AM, *et al.* South Carolina tuberculosis genotype cluster investigation: A tale of substance abuse and recurrent disease. *International Journal of Tuberculosis and Lung Disease* **14** 1347-1349. (2010)

240. McConkey SJ, *et al.* Prospective use of molecular typing of Mycobacterium tuberculosis by use of restriction fragment-length polymorphism in a public tuberculosis-control program. *Clinical Infectious Diseases* **34** 612-619. (2002)

241. Faccini M, *et al.* Tuberculosis-related stigma leading to an incomplete contact investigation in a low-incidence country. *Epidemiology and infection* **143** 2841. (2015)

242. Powell KM, *et al.* Outbreak of Drug-Resistant Mycobacterium tuberculosis Among Homeless People in Atlanta, Georgia, 2008-2015. *Public Health Rep* **132** 231-240. (2017)

243. Buskin SE, *et al.* Transmission cluster of multiclass highly drug-resistant HIV-1 among 9 men who have Sex with men in Seattle/King County, WA, 2005-2007. *Journal of Acquired Immune Deficiency Syndromes* **49** 205-211. (2008)

244. Hatzakis A, *et al.* Design and baseline findings of a large-scale rapid response to an HIV outbreak in people who inject drugs in Athens, Greece: the ARISTOTLE programme. *Addiction* **110** 1453-1467. (2015)

245. Moonan PK, *et al.* Using GIS technology to identify areas of tuberculosis transmission and incidence. *International Journal of Health Geographics [Electronic Resource]* **3** 13. (2004)

246. Moonan K, *et al.* What is the outcome of targeted tuberculosis screening based on universal genotyping and location? *American journal of respiratory and critical care medicine* **174** 599. (2006)

247. Ciccozzi M, *et al.* May phylogenetic analysis support epidemiological investigation in identifying the source of HIV infection? *AIDS Research and Human Retroviruses* **27** 455-457. (2011)

248. Munang ML, *et al.* Incorporating tuberculosis strain typing data into routine contact tracing investigations: Experience from the field. *Thorax* **68**. (2013)

249. Mitruka K, Oeltmann JE, Ijaz K, Haddad MB. Tuberculosis outbreak investigations in the United States, 2002-2008. *Emerg Infect Dis* **17** 425-431. (2011)

250. Dombrowski JC, *et al.* HIV provider and patient perspectives on the Development of a Health Department "Data to Care" Program: a qualitative study. *BMC Public Health* **16** 491. (2016)

251. Lee SS, *et al.* An exploratory study on the social and genotypic clustering of HIV infection in men having sex with men. *AIDS* **23** 1755-1764. (2009)

252. Oster AM, *et al.* Network analysis among HIV-infected young black men who have sex with men demonstrates high connectedness around few venues. *Sex Transm Dis* **40** 206-212. (2013)

253. Prost A, *et al.* "There is such a thing as asking for trouble": taking rapid HIV testing to gay venues is fraught with challenges. *Sex Transm Infect* **83** 185-188. (2007)

254. Mears J, *et al.* The prospective evaluation of the TB strain typing service in England: a mixed methods study. *Thorax* **71** 734-741. (2016)

255. Nicoll A, *et al.* Assessing the impact of national anti-HIV sexual health campaigns: trends in the transmission of HIV and other sexually transmitted infections in England. *Sex Transm Infect* **77** 242-247. (2001)

256. Brown AE, *et al.* Fall in new HIV diagnoses among men who have sex with men (MSM) at selected London sexual health clinics since early 2015: testing or treatment or pre-exposure prophylaxis (PrEP)? *Euro Surveill* **22**. (2017)

257. Practical guidelines for intensifying HIV prevention: towards universal access. Geneva: Joint United Nations Programme on HIV/AIDS; 2007. Available from http://data.unaids.org/pub/manual/2007/20070306_prevention_guidelines_towards_universal_access_en.pdf (Accessed June 2017)

258. Case KK, *et al.* Understanding the modes of transmission model of new HIV infection and its use in prevention planning. *Bull World Health Organ* **90** 831-838a. (2012)

259. Volz EM, *et al.* HIV-1 Transmission during Early Infection in Men Who Have Sex with Men: A Phylodynamic Analysis. *PLoS Med* **10**. (2013)

260. Ratmann O, *et al.* Sources of HIV infection among men having sex with men and implications for prevention. *Sci Transl Med* **8** 320ra322. (2016)

261. Volz EM, Frost SDW. Inferring the Source of Transmission with Phylogenetic Data. *PLoS Computational Biology* **9**. (2013)

262. Health Protection Agency. HIV in the United Kingdom: 2012 Report. London: Health Protection Services, Colindale. November 2012. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/335452/HIV_annual_report_2012.pdf (Accessed June 2018)

263. Cope AB, *et al.* Ongoing HIV transmission and the HIV care continuum in North Carolina. *PLoS ONE* **10**. (2015)

264. Marks G, Crepaz N, Janssen RS. Estimating sexual transmission of HIV from persons aware and unaware that they are infected with the virus in the USA. *Aids* **20** 1447-1450. (2006)

265. Hall HI, Holtgrave DR, Maulsby C. HIV transmission rates from persons living with HIV who are aware and unaware of their infection. *Aids* **26** 893-896. (2012)

266. Skarbinski J, *et al.* Human immunodeficiency virus transmission at each step of the care continuum in the United States. *JAMA Intern Med* **175** 588-596. (2015)

267. Johnson VA, *et al.* Update of the drug resistance mutations in HIV-1: March 2013. *Top Antivir Med* **21** 6-14. (2013)

268. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30** 1312-1313. (2014)

269. Shapiro B, Rambaut A, Drummond AJ. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Mol Biol Evol* **23** 7-9. (2006)

270. Minin VN, Bloomquist EW, Suchard MA. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol Biol Evol* **25** 1459-1471. (2008)

271. Birungi J, *et al.* Lack of Effectiveness of Antiretroviral Therapy in Preventing HIV Infection in Serodiscordant Couples in Uganda: An Observational Study. *PLoS One* **10** e0132182. (2015)

272. Hollingsworth TD, Anderson RM, Fraser C. HIV-1 transmission, by stage of infection. *J Infect Dis* **198** 687-693. (2008)

273. Pinkerton SD. How many sexually-acquired HIV infections in the USA are due to acute-phase HIV transmission? *Aids* **21** 1625-1629. (2007)

274. Xiridou M, Geskus R, de Wit J, Coutinho R, Kretzschmar M. Primary HIV infection as source of HIV transmission within steady and casual partnerships among homosexual men. *Aids* **18** 1311-1320. (2004)

275. Coutinho FA, Lopez LF, Burattini MN, Massad E. Modelling the natural history of HIV infection in individuals and its epidemiological implications. *Bull Math Biol* **63** 1041-1062. (2001)

276. Brown AE, *et al.* Phylogenetic reconstruction of transmission events from individuals with acute HIV infection: toward more-rigorous epidemiological definitions. *J Infect Dis* **199** 427-431. (2009)

277. Prah P, *et al.* Men who have sex with men in Great Britain: comparing methods and estimates from probability and convenience sample surveys. *Sex Transm Infect* **92** 455-463. (2016)

278. Rice BD, Elford J, Yin Z, Delpech VC. A new method to assign country of HIV infection among heterosexuals born abroad and diagnosed with HIV. *Aids* **26** 1961-1966. (2012)

279. Public health surveillance. Geneva: World Health Organization; 2017. Available from https://www.who.int/topics/public_health_surveillance/en/ (Accessed July 2018)

280. WHO guidelines on ethical issues in public health surveillance Geneva; 2007. vailable from https://www.who.int/ethics/publications/public-health-surveillance/en/ (Accessed February 2019)

281. Pneumocystis pneumonia--Los Angeles. *MMWR Morb Mortal Wkly Rep* **30** 250-252. (1981)

282. Guidelines for second generation HIV surveillance. An update: know your epidemic. *WHO*. (2014) Available from https://www.who.int/hiv/pub/surveillance/2013package/module1/en/ (Accessed July 2018)

283. Tostevin A, *et al.* Recent trends and patterns in HIV-1 transmitted drug resistance in the United Kingdom. *HIV Med* **18** 204-213. (2017)

284. International Health Regulations. 3rd Edition ed. Geneva; 2005.

285. Sexually transmitted infections (STIs): managing outbreaks. Public Health England, 2017. Available from: https://www.gov.uk/government/publications/sexually-transmitted-infections-stis-managing-outbreaks

286. Udeagu N, Webster R, Bocour A, Michel P, Shepard W. Lost or just not following up: public health effort to re-engage HIV-infected persons lost to follow-up into HIV medical care. *AIDS (London, England)* **27** 2271. (2013)

287. Buskin SE, Kent JB, Dombrowski JC, Golden MR. Migration distorts surveillance estimates of engagement in care: results of public health investigations of persons who appear to be out of HIV care. *Sex Transm Dis* **41** 35-40. (2014)

288. Bove JM, Golden MR, Dhanireddy S, Harrington RD, Dombrowski JC. Outcomes of a Clinic-Based Surveillance-Informed Intervention to Relink Patients to HIV Care. *J Acquir Immune Defic Syndr* **70** 262-268. (2015)

289. Herwehe J, *et al.* Implementation of an innovative, integrated electronic medical record (EMR) and public health information exchange for HIV/AIDS. *J Am Med Inform Assoc* **19** 448-452. (2012)

290. Buskin SE, *et al.* HIV infected individuals presumed to not be receiving HIV medical care: a surveillance program evaluation for investigations and referrals in Seattle, Washington State, USA. *Journal of HIV/AIDS Surveillance Epidemiology* **3** 3. (2011)

291. Bertolli J, *et al.* Missed Connections: HIV-Infected People Never in Care. *Public Health Rep* **128**, 117-126. (2013)

292. Padilla M, *et al.* Locating People Diagnosed With HIV for Public Health Action: Utility of HIV Case Surveillance and Other Data Sources. *Public Health Rep* **133** 147-154. (2018)

293. Fairchild AL, Bayer R. HIV surveillance, public health, and clinical medicine--will the walls come tumbling down? *N Engl J Med* **365** 685-687. (2011)

294. Evans D, *et al.* Acceptance of the use of HIV surveillance data for care engagement: national and local community perspectives. *J Acquir Immune Defic Syndr* **69 Suppl 1** S31-36. (2015)

295. Gordis L, Gold E, Seltser R. Privacy protection in epidemiologic and medical research: a challenge and a responsibility. *Am J Epidemiol* **105** 163-168. (1977)

296. Rothman KJ. The rise and fall of epidemiology, 1950--2000 A.D. *N Engl J Med* **304** 600-602. (1981)

297. 1991 International Guidelines for Ethical Review of Epidemiological Studies: Council for International Organizations of Medical Sciences; 1991.

298. International Health Regulations (1969). Geneva: World Health Organization; 1969.

299. Verity C, Nicoll A. Consent, confidentiality, and the threat to public health surveillance. *BMJ* **324** 1210-1213. (2002)

300. Public Health: Ethical Issues. London: Nuffield Council on Bioethics; November 2007.

301. Beauchamp TL, Childress JF. *Principles of Biomedical Ethics*. Oxford University Press: Oxford, 2001.

302. Last JM. *Dictionary of Public Health*, 1st edn. Oxford University Press: New York, 2007.

303. Sweeney P, *et al.* Shifting the paradigm: using HIV surveillance data as a foundation for improving HIV care and preventing HIV infection. *Milbank Q* **91** 558-603. (2013)

304. Managing Ethical Issues in Infectious Disease Outbreaks. Geneva: World Health Organization; 2016.

305. Fairchild AL, Bayer R, Colgrove J. *Searching Eyes: Privacy, the State, and Disease Surveillance in America*. University of California Press: Berkeley, 2007.

306. Bayer R. Public health policy and the AIDS epidemic. An end to HIV exceptionalism? *N Engl J Med* **324** 1500-1504. (1991)

307. Fisher WA, Kohut T, Fisher JD. AIDS Exceptionalism: On the Social Psychology of HIV Prevention Research. *Soc Issues Policy Rev* **3** 45-77. (2009)

308. Rosenbrock R, *et al.* The normalization of AIDS in Western European countries. *Soc Sci Med* **50** 1607-1629. (2000)

309. Rump B, Cornelis C, Woonink F, Verweij M. The need for ethical reflection on the use of molecular microbial characterisation in outbreak management. *Euro Surveill* **18** 20384. (2013)

310. Rump BO, Midden-Nederland MHSG, Woonink F, Midden-Nederland MHSG. Ethical Questions Concerning the Use of Molecular Typing Techniques in the Control of Infectious Diseases. *Public Health Ethics* **5** 311-313. (2019)

311. Bubela T, School of Public Health UoA, Yanow S, Health RaDAPLfP. Molecular Typing Technology: a Legal Perspective. *Public Health Ethics* **5** 317-320. (2012)

312. Emanuel EJ. The Oxford Textbook of Clinical Research Ethics. Oxford University Press, 2011.

313. Pillay D, *et al.* PANGEA-HIV: phylogenetics for generalised epidemics in Africa. *Lancet Infect Dis* **15** 259-261. (2015)

314. Coltart CEM, *et al.* Ethical considerations in global HIV phylogenetic research. *Lancet HIV* **5** e656-e666. (2018)

315. Mutenherwa F, Wassenaar DR, de Oliveira T. Experts' Perspectives on Key Ethical Issues Associated With HIV Phylogenetics as Applied in HIV Transmission Dynamics Research. *J Empir Res Hum Res Ethics* **14** 61-77. (2019)

316. Evans N. Ethical Considerations for a Public Health Response Using Molecular HIV Surveillance Data: A Multi-Stakeholder Approach: Project Inform and Northwestern University; February 2018.

317. Mehta SR, Schairer C, Little S. Ethical issues in HIV phylogenetics and molecular epidemiology. *Curr Opin HIV AIDS* **14** 221-226. (2019)

318. Wertheim JO, Kosakovsky Pond SL, Little SJ, De Gruttola V. Using HIV transmission networks to investigate community effects in HIV prevention trials. *PLoS One* **6** e27775. (2011)

319. Mehta SR, *et al.* Using HIV Sequence and Epidemiologic Data to Assess the Effect of Self-referral Testing for Acute HIV Infection on Incident Diagnoses in San Diego, California. *Clin Infect Dis* **63** 101-107. (2016)

320. Hughes GJ, *et al.* Molecular phylodynamics of the heterosexual HIV epidemic in the United Kingdom. *PLoS Pathog* **5** e1000590. (2009)

321. Hurt CB, Dennis AM. Putting it all together: lessons from the Jackson HIV outbreak investigation. *Sex Transm Dis*, vol. 40: United States, 2013, pp 213-215.

322. Burris S. Surveillance, social risk, and symbolism: framing the analysis for research and policy. *J Acquir Immune Defic Syndr* **25 Suppl 2** S120-127. (2000)

323. Gostin LO, Lazzarini Z, Neslund VS, Osterholm MT. The public health information infrastructure. A national review of the law on health information privacy. *Jama* **275** 1921-1927. (1996)

324. Bayer R, Levine C, Murray TH. Guidelines for confidentiality in research on AIDS. *Irb* **6** 1-7. (1984)

325. Mehta SR, Vinterbo SA, Little SJ. Ensuring privacy in the study of pathogen genetics. *Lancet Infect Dis* **14** 773-777. (2014)

326. Leitner T, *et al*. Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis. *Proc Natl Acad Sci USA* **93** 10864-9. (1996)

327. Resik S, *et al.* Limitations to contact tracing and phylogenetic analysis in establishing HIV type 1 transmission networks in Cuba. *AIDS Research and Human Retroviruses* **23** 347-356. (2007)

328. Cohen MS, McCauley M, Sugarman J. The Ethical Odyssey in Testing HIV Treatment as Prevention. *Clin Trials* **9** 340-347. (2012)

329. Offences Against the Person Act 1861. Available from http://www.legislation.gov.uk/ukpga/Vict/24-25/100 (Accessed 20 January 2019).

330. Bernard E, Azad Y, Delpech V, Geretti AM. HIV Forensics II: estimating the likelihood of recent HIV infection: implications for criminal prosecution. London: National AIDS Trust; 2011.

331. Intentional or Reckless Sexual Transmission of Infection. The Crown Prosecution Service. 2007. Available from: https://www.cps.gov.uk/legal-guidance/intentional-or-reckless-sexual-transmission-infection (Accessed February 2019)

332. Abecasis AB, Pingarilho M, Vandamme AM. Phylogenetic analysis as a forensic tool in HIV transmission investigations. *Aids* **32** 543-554. (2018)

333. Police and Criminal Evidence Act 1984. Available from http://www.legislation.gov.uk/ukpga/1984/60/contents (Accessed March 2019).

334. Dyer C. Use of confidential HIV data helps convict former prisoner. *Bmj* **322** 633. (2001)

335. Phillips MD, Schembri G. Narratives of HIV: measuring understanding of HIV and the law in HIV-positive patients. *J Fam Plann Reprod Health Care* **42** 30-35. (2016)

336. Dodds C, Bourne A, Weait M. Responses to criminal prosecutions for HIV transmission among gay men with HIV in England and Wales. *Reprod Health Matters* **17** 135-145. (2009)

337. General Medical Council. Confidentiality: Good Practice in Handling Patient Information. 2017. Available from https://www.gmc-uk.org/ethical-guidance/ethical-guidance-for-doctors/confidentiality. (Accessed March 2019).

338. Schairer CE, *et al.* Trust and Expectations of Researchers and Public Health Departments for the Use of HIV Molecular Epidemiology. *AJOB Empir Bioeth* 1-13. (2019)

339. Burris S, Gable L, Stone L, Lazzarini Z. The role of state law in protecting human subjects of public health research and practice. *J Law Med Ethics* **31** 654-662. (2003)

340. Kass NE. An ethics framework for public health. *Am J Public Health* **91** 1776-1782. (2001)

341. Davies A, Scott S, Badger S, Torok ME, Peacock SJ. Public perceptions of bacterial whole-genome sequencing for tuberculosis. *Trends in Genetics* **31** 58-60. (2015)

342. Yebra G, *et al.* Phylogenetic Analysis of HIV Full Genomes in London, UK: Initial Results From ICONIC. Presented at: Conference on Retroviruses and Opportunistic Infections 2016; February 22-25, 2016; Boston, MA

343. Rachinger A, Groeneveld PHP, Van S, Lemey P, Schuitemaker H. Time-measured phylogenies of gag, pol and env sequence data reveal the direction and time interval of HIV-1 transmission. *AIDS* **25** 1035-1039. (2011)

344. Romero-Severson E. Sex, drugs, and phylogeny: The ABC's of a scientific soap opera. Session presented at: 24th International HIV Dynamics & Evolution; 2017; Isle of Skye, Scotland; 2017.

345. Crotty M. *The Foundations of Social Research: Meaning and Perspective in the Research Process*. SAGE Publications, 1998.

346. Dewey J. *Logic - The Theory of Inquiry*. Read Books, 2008.

347. Niglas K. The Multidimensional Model of Research Methodology: An Integrated Set of Continua. In: Tashakkori A, Teddlie C (eds). *SAGE Handbook of Mixed Methods in Social & Behavioral Research*. SAGE Publications, Inc.: Thousand Oaks, 2015, pp 215-236.

348. Braun V, Clarke V. Using thematic analysis in psychology. *Qualitative Research in Psychology* **3** 77-101. (2006)

349. Goodyear-Smith F, Jackson C, Greenhalgh T. Co-design and implementation research: challenges and solutions for ethics committees. *BMC Med Ethics* **16** 78. (2015)

350. Sanders EBN, Stappers PJ. Co-creation and the new landscapes of design. *CoDesign* **4** 5-18. (2008)

351. Pao D, *et al.* Transmission of HIV-1 during primary infection: Relationship to sexual risk and sexually transmitted infections. *AIDS* **19** 85-90. (2005)

352. Woith WM, Jenkins SH, Astroth KS, Kennedy JA. Lessons learned from conducting qualitative research in a hospital. *Nurse Res* **22** 40-43. (2014)

353. Bird SM, Brown AJ. Criminalisation of HIV transmission: implications for public health in Scotland. *Bmj* **323** 1174-1177. (2001)

354. Lee LM, Heilig CM, White A. Ethical Justification for Conducting Public Health Surveillance Without Patient Consent. *Am J Public Health* **102** 38-44. (2012)

355. Hodcroft EB. Estimating the heritability of virulence in HIV. PhD thesis, University of Edinburgh, 2015.

8. Appendices

Appendix A. Ethics approval for phylogenetic study (chapter 4)

Appendix B. Ethics approval for qualitative study (chapter 5)

Appendix C. BHIVA Research Award funding letter for qualitative study (chapter 5)

Appendix D. Advertising flyers and posters for qualitative study (chapter 5)

Appendix E. Patient information and consent forms used for qualitative study (chapter 5)

Appendix F. Demographic questionnaire used for qualitative study (chapter 5)

Appendix G. PowerPoint presentations to illustrate concepts and scenarios for qualitative study (chapter 5)

Appendix H. Post discussion information sheet provided to participants in qualitative study (chapter 5)

Appendix A. Ethics approval for phylogenetic study (chapter 4)

NHS
Health Research Authority
National Research Ethics Service

London - Queen Square Research Ethics Committee
HRA NRES Centre Manchester
Barlow House
3rd Floor
4 Minshull Street
Manchester
M1 3DZ

16 March 2016

Dr Larissa Mulka
Room 318b Mayfield House,
University of Brighton
Falmer
BN1 9PH

Dear Dr Mulka

Study title:	**A combined clinical, epidemiological and phylogenetic analysis of the Brighton HIV cohort to assess the current pattern of transmission networking, identify factors associated with transmission and determine the proportion who have acquired infection from an outside or undiagnosed source.**
REC reference:	**16/LO/0539**
Protocol number:	**195958**
IRAS project ID:	**195958**

The Proportionate Review Sub-committee of the London - Queen Square Research Ethics Committee reviewed the above application on 17 March 2016.

We plan to publish your research summary wording for the above study on the HRA website, together with your contact details. Publication will be no earlier than three months from the date of this favourable opinion letter. The expectation is that this information will be published for all studies that receive an ethical opinion but should you wish to provide a substitute contact point, wish to make a request to defer, or require further information, please contact the REC Manager Rachel Heron, nrescommittee.london-queensquare@nhs.net Under very limited circumstances (e.g. for student research which has received an unfavourable opinion), it may be possible to grant an exemption to the publication of the study.

Ethical opinion

On behalf of the Committee, the sub-committee gave a favourable ethical opinion of the above research on the basis described in the application form, protocol and supporting documentation, subject to the conditions specified below.

Conditions of the favourable opinion

The REC favourable opinion is subject to the following conditions being met prior to the start

of the study.

Management permission must be obtained from each host organisation prior to the start of the study at the site concerned.

Management permission should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements. Each NHS organisation must confirm through the signing of agreements and/or other documents that it has given permission for the research to proceed (except where explicitly specified otherwise).

Guidance on applying for HRA Approval (England)/ NHS permission for research is available in the Integrated Research Application System, www.hra.nhs.uk or at http://www.rdforum.nhs.uk.

Where a NHS organisation's role in the study is limited to identifying and referring potential participants to research sites ("participant identification centre"), guidance should be sought from the R&D office on the information it requires to give permission for this activity.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of management permissions from host organisations.

Registration of Clinical Trials

All clinical trials (defined as the first four categories on the IRAS filter page) must be registered on a publically accessible database. This should be before the first participant is recruited but no later than 6 weeks after recruitment of the first participant.

There is no requirement to separately notify the REC but you should do so at the earliest opportunity e.g. when submitting an amendment. We will audit the registration details as part of the annual progress reporting process.

To ensure transparency in research, we strongly recommend that all research is registered but for non-clinical trials this is not currently mandatory.

If a sponsor wishes to request a deferral for study registration within the required timeframe, they should contact hra.studyregistration@nhs.net. The expectation is that all clinical trials will be registered, however, in exceptional circumstances non registration may be permissible with prior agreement from the HRA. Guidance on where to register is provided on the HRA website.

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

Ethical review of research sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion").

Approved documents

The documents reviewed and approved were:

Document	*Version*	*Date*
Covering letter on headed paper [Cover letter]	1.0	02 March 2016
IRAS Checklist XML [Checklist_08032016]		08 March 2016
REC Application Form [REC_Form_08032016]		08 March 2016
Research protocol or project proposal [Protocol]	1.1	22 February 2016
Summary CV for Chief Investigator (CI) [Summary CV L Mulka]	1.0	02 March 2016
Summary CV for supervisor (student research) [Supervisor CV - J Vera]	1.0	01 July 2015

Membership of the Proportionate Review Sub-Committee

The members of the Sub-Committee who took part in the review are listed on the attached sheet.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Reporting requirements

The attached document "After ethical review – guidance for researchers" gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Notification of serious breaches of the protocol
- Progress and safety reports
- Notifying the end of the study

The HRA website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

User Feedback

The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application procedure. If you wish to make your views known please use the feedback form available on the HRA website: http://www.hra.nhs.uk/about-the-hra/governance/quality-assurance/

HRA Training

We are pleased to welcome researchers and R&D staff at our training days – see details at http://www.hra.nhs.uk/hra-training/

With the Committee's best wishes for the success of this project.

16/LO/0539	Please quote this number on all correspondence

Yours sincerely

Signed on behalf of
Dr Simon Eaton
Chair

Email: nrescommittee.london-queensquare@nhs.net

Enclosures: *List of names and professions of members who took part in the review*

"After ethical review – guidance for researchers" [SL-AR2]

Copy to: *Dr Larissa Mulka*
Dr Scott Harfield, Brighton & Sussex University Hospitals NHS Trust

Appendix B. Ethics approval for qualitative study (chapter 5)

NHS
Health Research Authority

Dr Larissa Mulka
Room 318b Mayfield House
University of Brighton
Falmer
BN1 9PH
l.mulka@bsms.ac.uk

Email: hra.approval@nhs.net

14 December 2017

Dear Dr Mulka,

Letter of HRA Approval

Study title:	**A rapid ethical assessment of the use of phylogenetics for public health interventions in HIV**
IRAS project ID:	**228210**
REC reference:	**17/LO/1795**
Sponsor	**Brighton and Sussex Universities NHS Hospitals Trust**

I am pleased to confirm that **HRA Approval** has been given for the above referenced study, on the basis described in the application form, protocol, supporting documentation and any clarifications noted in this letter.

Participation of NHS Organisations in England

The sponsor should now provide a copy of this letter to all participating NHS organisations in England.

Appendix B provides important information for sponsors and participating NHS organisations in England for arranging and confirming capacity and capability. **Please read *Appendix B* carefully**, in particular the following sections:

- *Participating NHS organisations in England* – this clarifies the types of participating organisations in the study and whether or not all organisations will be undertaking the same activities
- *Confirmation of capacity and capability* - this confirms whether or not each type of participating NHS organisation in England is expected to give formal confirmation of capacity and capability. Where formal confirmation is not expected, the section also provides details on the time limit given to participating organisations to opt out of the study, or request additional time, before their participation is assumed.
- *Allocation of responsibilities and rights are agreed and documented (4.1 of HRA assessment criteria)* - this provides detail on the form of agreement to be used in the study to confirm capacity and capability, where applicable.

Further information on funding, HR processes, and compliance with HRA criteria and standards is also provided.

IRAS project ID	228210

It is critical that you involve both the research management function (e.g. R&D office) supporting each organisation and the local research team (where there is one) in setting up your study. Contact details and further information about working with the research management function for each organisation can be accessed from www.hra.nhs.uk/hra-approval.

Appendices

The HRA Approval letter contains the following appendices:

- A – List of documents reviewed during HRA assessment
- B – Summary of HRA assessment

After HRA Approval

The document *"After Ethical Review – guidance for sponsors and investigators"*, issued with your REC favourable opinion, gives detailed guidance on reporting expectations for studies, including:

- Registration of research
- Notifying amendments
- Notifying the end of the study

The HRA website also provides guidance on these topics, and is updated in the light of changes in reporting expectations or procedures.

In addition to the guidance in the above, please note the following:

- HRA Approval applies for the duration of your REC favourable opinion, unless otherwise notified in writing by the HRA.
- Substantial amendments should be submitted directly to the Research Ethics Committee, as detailed in the *After Ethical Review* document. Non-substantial amendments should be submitted for review by the HRA using the form provided on the HRA website, and emailed to hra.amendments@nhs.net.
- The HRA will categorise amendments (substantial and non-substantial) and issue confirmation of continued HRA Approval. Further details can be found on the HRA website.

Scope

HRA Approval provides an approval for research involving patients or staff in NHS organisations in England.

If your study involves NHS organisations in other countries in the UK, please contact the relevant national coordinating functions for support and advice. Further information can be found at http://www.hra.nhs.uk/resources/applying-for-reviews/nhs-hsc-rd-review/.

If there are participating non-NHS organisations, local agreement should be obtained in accordance with the procedures of the local participating non-NHS organisation.

User Feedback

The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application

IRAS project ID	228210

procedure. If you wish to make your views known please use the feedback form available on the HRA website: http://www.hra.nhs.uk/about-the-hra/governance/quality-assurance/.

HRA Training

We are pleased to welcome researchers and research management staff at our training days – see details at http://www.hra.nhs.uk/hra-training/

Your IRAS project ID is **228210**. Please quote this on all correspondence.

Yours sincerely

Gemma Oakes
Assessor

Email: hra.approval@nhs.net

Copy to: *Mr Scott Harfield, Brighton And Sussex University Hospitals NHS Trust [Sponsor Contact & Lead NHS R&D Contact]*
r&d.approvals@bsuh.nhs.uk

IRAS project ID	228210

Appendix A - List of Documents

The final document set assessed and approved by HRA Approval is listed below.

Document	*Version*	*Date*
Copies of advertisement materials for research participants [Recruitment advert for men who have sex with men]	1.1	04 September 2017
Copies of advertisement materials for research participants [Recruitment advert for black African men and women]	1.1	04 September 2017
Copies of advertisement materials for research participants [Clinic poster to alert participants to observational study]	1.0	14 November 2017
Interview schedules or topic guides for participants [Slides and crib sheets for HIV positive participants]	1.1	04 September 2017
Interview schedules or topic guides for participants [Slides and crib sheets for HIV negative participants]	1.1	04 September 2017
Interview schedules or topic guides for participants [Slides and crib sheets for Healthcare worker participants]	1.1	04 September 2017
IRAS Application Form [IRAS_Form_02102017]		02 October 2017
IRAS Application Form XML file [IRAS_Form_02102017]		02 October 2017
IRAS Checklist XML [Checklist_06102017]		06 October 2017
Letter from funder		12 October 2017
Letter from funder [BHIVA Research Award letter]	1.0	09 August 2017
Letter from sponsor [Letter of approval from sponsor]	1.0	25 September 2017
Letters of invitation to participant [Flyer for potential observation participants]	1.0	14 November 2017
Non-validated questionnaire [Demographic questions for HIV positive participants]	1.1	04 September 2017
Non-validated questionnaire [Demographic Questions for HIV-negative participants]	1.1	04 September 2017
Participant information sheet (PIS) [Post-interview information leaflet]	1.1	04 September 2017
Participant information sheet (PIS) [PIS and consent form for partner notification observation (Healthcare workers) with tracked changes]	2.0	14 November 2017
Participant information sheet (PIS) [PIS and consent form for partner notification observation (Healthcare workers) tracked changes removed]	2.0	14 November 2017
Participant information sheet (PIS) [PIS and consent form for partner notification observation (non-Healthcare workers) with tracked changes]	2.0	14 November 2017
Participant information sheet (PIS) [PIS and consent form for partner notification observation (non-Healthcare workers) tracked changes removed]	2.0	14 November 2017
Participant information sheet (PIS) [PIS and consent form for interviews and focus group discussions (Healthcare workers) with tracked changes]	2.0	14 November 2017
Participant information sheet (PIS) [PIS and consent form for interviews and focus group discussions (Healthcare workers) tracked changes removed]	2.0	14 November 2017
Participant information sheet (PIS) [PIS and consent form for interviews and focus group discussions (non-Healthcare workers) with tracked changes]	2.0	14 November 2017
Participant information sheet (PIS) [PIS and consent form for interviews and focus group discussions (non-Healthcare workers) tracked changes removed]	2.0	14 November 2017

IRAS project ID	228210

Participant information sheet (PIS) [PIS and consent form for blood-taking observation (non-Healthcare workers) with tracked changes]	2.0	14 November 2017
Participant information sheet (PIS) [PIS and consent form for blood-taking observation (non-Healthcare workers) tracked changes removed]	2.0	14 November 2017
Participant information sheet (PIS) [PIS and consent form for blood-taking observation (Healthcare workers) with tracked changes]	2.0	14 November 2017
Participant information sheet (PIS) [PIS and consent form for blood-taking observation (Healthcare workers) tracked changes removed]	2.0	14 November 2017
Research protocol or project proposal [Protocol]	1.1	04 September 2017
Summary CV for Chief Investigator (CI) [CV for Chief Investigator - L Mulka]	1.0	11 September 2017
Summary CV for student [Student CV - Dr L Mulka]	1.0	11 September 2017
Summary CV for supervisor (student research) [Supervisor CV - Prof. J Cassell]	1.0	11 September 2017

IRAS project ID	228210

Appendix B - Summary of HRA Assessment

This appendix provides assurance to you, the sponsor and the NHS in England that the study, as reviewed for HRA Approval, is compliant with relevant standards. It also provides information and clarification, where appropriate, to participating NHS organisations in England to assist in assessing and arranging capacity and capability.

For information on how the sponsor should be working with participating NHS organisations in England, please refer to the *participating NHS organisations, capacity and capability* and *Allocation of responsibilities and rights are agreed and documented (4.1 of HRA assessment criteria)* sections in this appendix.

The following person is the sponsor contact for the purpose of addressing participating organisation questions relating to the study:

Name: Mr Scott Harfield
Tel: 01273 696 955 (Ext: 7497)
Email: r&d.approvals@bsuh.nhs.uk

HRA assessment criteria

Section	HRA Assessment Criteria	Compliant with Standards	Comments
1.1	IRAS application completed correctly	Yes	No comments.
2.1	Participant information/consent documents and consent process	Yes	The applicant has confirmed that only members of the routine care team will access medical notes to identify potential participants.
3.1	Protocol assessment	Yes	No comments.
4.1	Allocation of responsibilities and rights are agreed and documented	Yes	This is a non-commercial single site study taking place in the NHS where that single NHS organisation is also the study sponsor. Therefore no study agreements are required.
4.2	Insurance/indemnity arrangements assessed	Yes	Where applicable, independent contractors (e.g. General Practitioners) should ensure that the professional indemnity provided by their medical defence organisation covers the

IRAS project ID	228210

Section	HRA Assessment Criteria	Compliant with Standards	Comments
			activities expected of them for this research study
4.3	Financial arrangements assessed	Yes	External study funding has been secured from BHIVA to run the study at site.
5.1	Compliance with the Data Protection Act and data security issues assessed	Yes	The applicant has confirmed that a confidentiality agreement will be entered into with Essential Secretary.
5.2	CTIMPS – Arrangements for compliance with the Clinical Trials Regulations assessed	Not Applicable	No comments
5.3	Compliance with any applicable laws or regulations	Yes	No comments
6.1	NHS Research Ethics Committee favourable opinion received for applicable studies	Yes	REC Favourable Opinion was issued on 06 December 2017.
6.2	CTIMPS – Clinical Trials Authorisation (CTA) letter received	Not Applicable	No comments
6.3	Devices – MHRA notice of no objection received	Not Applicable	No comments
6.4	Other regulatory approvals and authorisations received	Not Applicable	No comments

Participating NHS Organisations in England

This provides detail on the types of participating NHS organisations in the study and a statement as to whether the activities at all organisations are the same or different.
This is a non-commercial single site study taking place in the NHS where that single NHS organisation is also the study sponsor. Therefore there is only one site type involved in the research. If this study is subsequently extended to other NHS organisation(s) in England, an amendment should be submitted to the HRA, with a Statement of Activities and Schedule of Events for the newly participating NHS organisation(s) in England. Please note that the remit of HRA Approval is limited to the NHS involvement in the study. Research

activity undertaken at non-NHS sites or involving non-NHS sites is therefore not covered and the research team should make appropriate alternative arrangements with relevant management at these organisations to conduct the research there.

The Chief Investigator or sponsor should share relevant study documents with participating NHS organisations in England in order to put arrangements in place to deliver the study. The documents should be sent to both the local study team, where applicable, and the office providing the research management function at the participating organisation. For NIHR CRN Portfolio studies, the Local LCRN contact should also be copied into this correspondence. For further guidance on working with participating NHS organisations please see the HRA website.

If chief investigators, sponsors or principal investigators are asked to complete site level forms for participating NHS organisations in England which are not provided in IRAS or on the HRA website, the chief investigator, sponsor or principal investigator should notify the HRA immediately at hra.approval@nhs.net. The HRA will work with these organisations to achieve a consistent approach to information provision.

Confirmation of Capacity and Capability

This describes whether formal confirmation of capacity and capability is expected from participating NHS organisations in England.
This is a single site study sponsored by the site. The R&D office will confirm to the CI when the study can start.

Principal Investigator Suitability

This confirms whether the sponsor position on whether a PI, LC or neither should be in place is correct for each type of participating NHS organisation in England and the minimum expectations for education, training and experience that PIs should meet (where applicable).
The sponsor has confirmed the Chief Investigator will act as the Local Principal Investigator at site. If this study is subsequently extended to other NHS organisation(s) in England, a further review of this section will be required. GCP training is not a generic training expectation, in line with the HRA statement on training expectations.

HR Good Practice Resource Pack Expectations

This confirms the HR Good Practice Resource Pack expectations for the study and the pre-engagement checks that should and should not be undertaken
In accordance with HR Good Practice Guidelines, local staff undertaking research activities at participating NHS sites will not require access arrangements. Where arrangements are not already in place, network staff (or similar) undertaking any research activities that do not impact on the quality of care of the participant (administering questionnaires, focus groups, interviews and observations), would be expected to obtain an honorary research

IRAS project ID	228210

contract from one NHS organisation (if university employed), followed by Letters of Access for subsequent organisations. This would be on the basis of a Research Passport (if university employed) or an NHS to NHS confirmation of pre-engagement checks letter (if NHS employed). These should confirm standard DBS checks and occupational health clearance.

Other Information to Aid Study Set-up

This details any other information that may be helpful to sponsors and participating NHS organisations in England to aid study set-up.

The applicant has indicated that they intend to apply for inclusion on the NIHR CRN Portfolio.

Appendix C. BHIVA Research Award funding letter for qualitative study (chapter 5)

British HIV Association
BHIVA

Secretariat: Mediscript Ltd
1 Mountview Court · 310 Friern Barnet Lane · London N20 0LD
Tel: +44 (0)20 8369 5380 · Fax: +44 (0)20 8446 9194 · Email: bhiva@bhiva.org · Web: www.bhiva.org

Patron: The Rt Hon Lord Fowler

9 August 2017

Dr Larissa Mulka
Brighton & Sussex Medical School
Mayfield House
Falmer
Brighton
BN1 9PH

By email to: l.mulka@bsms.ac.uk

Dear Dr Mulka

BHIVA Research Awards 2017

Award Type: Other
Project title: What is the acceptability of using phylogenetic data in clinical and public health practice? A Qualitative study aimed at intervention development

The BHIVA Research Awards Judging Panel met on 27 June 2017 under the leadership of the Panel, and considered the applications for the BHIVA Research Awards 2017. This year we received 16 applications and the standard was extremely high. The aggregate value of awards BHIVA has been approved and will be approximately £60k to successful applicants. Details of the winners will be posted on the BHIVA website at www.bhiva.org in due course.

We are pleased to confirm that your application for funding the above project has been successful and an award of £3,269 has been granted. Many congratulations on behalf of the Panel. We note that your project is due to commence on 1 December 2017 with a completion date of 31 March 2018.

Full details of the conditions of your award can be found in the BHIVA Research Award Guidance Manual. Please see below, some of the award requirements:

- Completed starting certificate form should be returned to the Secretariat on commencement of project.
- Interim mid-project report will be due half way into the 4 month project timeline.
- Final report will be due within 4 weeks of completion of your project. The report form will be issued by the Secretariat.
- You are required to submit an abstract or a plenary presentation to the BHIVA Annual Conference (abstract/poster to include BHIVA logo and appropriate acknowledgement of support).
- You are required to identify how BHIVA will be acknowledged for funding your project in your final report, including details of journal publications or submissions following completion of your project

Please note that 10% of the grant will be retained by BHIVA and released upon submission of your final report and abstract.

Kind regards

Yours sincerely

Dr Ian Williams
Chair, BHIVA Research Awards Judging Panel

Prof Caroline Sabin
Chair, BHIVA Education and Scientific Subcommittee

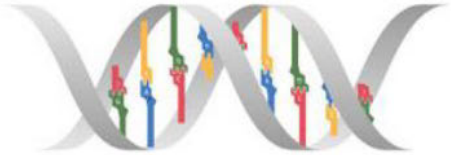
cc: Finance Officer of organisation
Enc: Starting certificate form

The BHIVA Research Awards 2017 are part-funded by Gilead Sciences. BHIVA is also grateful for donations from BHIVA members.

Registered Charity No. 1170707 · VAT Registration No. 689 5177 69

Appendix D. Advertising flyers and posters for qualitative study (chapter 5)

V 1.1, 04.09.2017. L Mulka



Should we use HIV genetic code to help diagnose the undiagnosed? Your views needed

Are you a black African man or woman living with HIV? Would you be willing to participate in one-on-one interview or a group discussion?

Interviews and separate group discussions for men and women will be held. We will try and arrange a time and location convenient for you.

No prior knowledge or experience is required

A £20 'thank you' is offered to everyone who takes part*

Your contributions in the focus group discussion will be recorded but kept anonymous. You will need to give your name and contact details for you to take part.

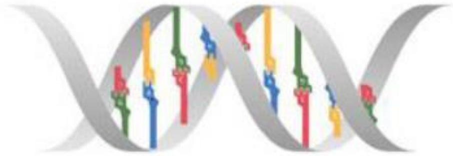
To reserve a place or for more information please contact Larissa Mulka at Brighton & Sussex Medical School: 01273 642185 larissa.mulka@bsuh.nhs.uk

Thank you

***This type of payment will not interfere with any benefit payments if you have not been involved in research with Brighton & Sussex Medical School in the past financial year**

This is a research project from Brighton and Sussex University Hospitals NHS Trust, funded by a BHIVA research award. Principal Investigator Dr Larissa Mulka.

V 1.1, 04.09.2017. L Mulka



Should we use HIV genetic code to help diagnose the undiagnosed? Your views needed

Are you a gay or bisexual man?
Would you be willing to participate in a group discussion or one-on-one interview?

Interviews and separate group discussions for HIV positive and negative men will be held at the Elton John Centre, RSCH
We will try to arrange a date and time that is convenient for you

No prior knowledge or experience is required

A £20 'thank you' is offered to everyone who takes part*

Your contributions in the focus group discussion will be recorded but kept anonymous.
We will need your name and contact details for you to take part.

To reserve a place or for more information please contact Larissa Mulka at
Brighton & Sussex Medical School: 01273 642185 larissa.mulka@bsuh.nhs.uk

Thank you

*This type of payment will not interfere with any benefit payments if you have not been involved in research with Brighton & Sussex Medical School in the past financial year

This is a research project from Brighton and Sussex University Hospitals NHS Trust, funded by a BHIVA research award.
Principal Investigator Dr Larissa Mulka. Email larissa.mulka@bsuh.nhs.uk

Discussion Group 2
Larissa Mulka
(01273) 642185
larissa.mulka@bsuh.nhs.uk

Discussion Group 2
Larissa Mulka
(01273) 642185
larissa.mulka@bsuh.nhs.uk

Discussion Group 2
Larissa Mulka
(01273) 642185
larissa.mulka@bsuh.nhs.uk

Discussion Group 2
Larissa Mulka
(01273) 642185
larissa.mulka@bsuh.nhs.uk

Discussion Group 2
Larissa Mulka
(01273) 642185
larissa.mulka@bsuh.nhs.uk

Discussion Group 2
Larissa Mulka
(01273) 642185
larissa.mulka@bsuh.nhs.uk

Discussion Group 2
Larissa Mulka
(01273) 642185
larissa.mulka@bsuh.nhs.uk

Appendix E. Patient information and consent forms used for qualitative study (chapter 5)

HIV/GUM Research Department, BSUH NHS Trust
The Elton John Centre
Sussex House
1 Abbey Road
BRIGHTON
BN2 1ES

Tel: 01273 523079

Fax: 01273 523080

Brighton and Sussex University Hospitals NHS Trust

Participant information sheet– Healthcare-worker Focus Group Discussion

Study title: A rapid ethical assessment of the use of phylogenetics for public health interventions in HIV.

You are being invited to take part in a research study. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this.

1. What is the purpose of the study?

A new way of using HIV data has been developed that could help us guide public health interventions in the future. This potential new method is using information taken from the HIV virus' genetic code.

This method is not currently being used, but we are collecting data to find out how acceptable using peoples HIV genetic code data this way might be to patients and healthcare staff.

HIV viruses in different people are very slightly different, and blood tests are taken routinely from everyone diagnosed with HIV to look for these differences, found in the genetic code of the virus. This is because the genetic code of the virus tells us what drugs work best against it.

Different people's HIV virus genetic codes can be compared to see how similar they are. We can perform analyses to find groups of similar viruses from different people which can help us understand patterns of HIV transmission. These analyses could be used in the future to help guide interventions to reduce the spread of HIV.

Before exploring how we could use people's information in this way, we need to find out what people understand about the processes that might be involved, and how they might feel about HIV virus information being used for this purpose, to ensure any future research in this area is acceptable to them.

2. Why have I been chosen?

You are being invited to take part in this study because you work as a healthcare professional with patients living with HIV.

3. Do I have to take part?

It is up to you whether or not to take part. If you decide to take part, you will be given this information sheet to keep and be asked to sign a consent form. A decision not to take part will not affect in any way how you are treated as a professional and does not exclude you from any studies in the future.

.

4. What will happen to me if I take part?

If you agree to join the study and sign the consent form at the back of this leaflet, you will be asked to join a focus group where you will be asked to discuss your thoughts, including the positives and negatives, of the use of phylogenetics for public health interventions and contact tracing in HIV, after being given some information and prompts. This will take up to 90 minutes, and be digitally voice recorded. You are not expected to have any prior knowledge on this subject, and anything you are unsure about will be explained to you by the researcher.

5. What are the possible disadvantages and risks of taking part?

We do not expect any direct risks to you from participating in this study, however this discussion will take up to 90 minutes of your time.

6. What if something goes wrong?

If you are harmed by taking part in this research project, there are no special compensation arrangements. If you are harmed due to someone's negligence, then you may have grounds for legal action but you may have to pay for it. Regardless of this, if you wish to complain, or have any concerns about any aspect of the way you have been approached or treated during this study, please contact Larissa Mulka, investigator of the study (larissa.mulka@bsuh.nhs.uk). If you remain unhappy and wish to complain formally, you can do this by contacting Brighton Patient Advice and Liaison Service on (01273) 664511 or via email – pals@bsuh.nhs.uk.

7. What are the possible benefits of taking part?

There are no benefits to you personally. Your participation will help us determine how technologies such as phylogenetics may be used in the future in a way that is acceptable to patients and staff, ensuring any concerns raised by this research are addressed and no interventions are introduced which may cause harm or distress.

8. Will my taking part in this study be kept confidential?

Confidentiality and anonymity will be highly protected, as follows:

The only identifiable information taken from you will be your job role (e.g. healthcare assistant, doctor). This will be kept in a locked file, only the researcher has access to. All information taken from focus groups, including your questionnaire data will be collected under a study ID from which you will not be identifiable. Digital recordings will be typed up, then securely destroyed. The responses you give during the discussion may be directly quoted in the study write up, but this will not be linked to any information which could reveal your identity.

9. What will happen to the results of the research study?

The results of the study will be presented at a medical conference and published in a medical journal and may be used to guide further research. Any quotes that may be included in this will not be linked to any data which might make you identifiable to the public.

10. What will happen to my data?

Your data (job title and transcribed focus group data) will be kept in separate password protected files on a secure University network for 5 years, as is routine in medical studies. Any paper copies will be kept in a secure locked cupboard for 5 years, then all data will be safely destroyed.

10. Who is organising and funding the research?

This study is being funded by the British HIV Association and is sponsored by Brighton & Sussex University Hospitals NHS Trust.

11. Contact for further information

You may ask questions about this study at any time. If you have any questions about the informed consent process or you require any further information relating to the study procedures or interviews, please contact Dr Larissa Mulka who is the researcher for this project. Email: Larissa.mulka@bsuh.nhs.uk

HIV/GUM Research Department, BSUH NHS Trust
The Elton John Centre
Sussex House
1 Abbey Road
BRIGHTON
BN2 1ES

Tel: 01273 523079

Fax: 01273 523080

Brighton and Sussex University Hospitals NHS Trust

CONSENT FORM

Title of project: A rapid ethical assessment of the use of phylogenetics for public health interventions in HIV.

Patient identification Number for this trial:

Names of investigator: Dr Larissa Mulka

Please initial box

1. I confirm that I have read and understand the information sheet dated 14.11.2017 for the above study and have had the opportunity to ask questions. ☐
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my professional rights being affected. ☐
3. I understand that my data will be digitally recorded then typed up. Things I say may be taken from the discussion and be directly quoted in the study write up, but no identifiable information will be included in this, and I will have the opportunity to review my quotes prior to publication. ☐
4. I agree to take part in the above study. ☐

Name of patient	Date	Signature
Name of person receiving consent (if different from researcher)	Date	Signature
Researcher	Date	Signature

1 copy for participant, 1 copy for site file

HIV/GUM Research Department, BSUH NHS Trust
The Elton John Centre
Sussex House
1 Abbey Road
BRIGHTON
BN2 1ES

Tel: 01273 523079

Fax: 01273 523080

Brighton and Sussex University Hospitals NHS Trust

Participant information sheet - Interviews and Focus Groups

Study title: A rapid ethical assessment of the use of phylogenetics for public health interventions in HIV.

You are being invited to take part in a research study. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this.

1. What is the purpose of the study?

A new way of using HIV data has been developed that could help us guide public health interventions in the future. This potential new method is using information taken from the HIV virus' genetic code.

This method is not currently being used, but we are collecting data to find out how acceptable using peoples HIV genetic code data this way might be to patients and healthcare staff.

HIV viruses in different people are very slightly different, and blood tests are taken routinely from everyone diagnosed with HIV to look for these differences, found in the genetic code of the virus. This is because the genetic code of the virus tells us what drugs work best against it.

Different people's HIV virus genetic codes can be compared to see how similar they are. We can perform analyses to find groups of similar viruses from different people which can help us understand patterns of HIV transmission. These analyses could be used in the future to help guide interventions to reduce the spread of HIV.

Before exploring how we could use people's information in this way, we need to find out what people understand about the processes that might be involved, and how they might feel about HIV virus information being used for this purpose, to ensure any future research in this area is acceptable to them.

2. Why have I been chosen?

You are being invited to take part in this study because you are HIV positive, or a HIV negative gay or bisexual man.

3. Do I have to take part?

It is up to you whether or not to take part. If you decide to take part, you will be given this information sheet to keep and sign a consent form. A decision not to take part will not affect in any way the standard of care you receive. It does not mean you cannot be involved in other studies in the future.

4. What will happen to me if I take part?

You will be asked to complete a brief questionnaire with details including sex, ethnicity and risk factor for HIV (i.e. sex with men, sex with women, injecting drugs).

You will either take part in a one-on-one interview with a researcher or join a focus group where you will be asked to discuss your thoughts on the use of 'phylogenetics', after being given some information and prompts. This process will take approximately 90 minutes, and will be digitally voice recorded.

You will receive £20 on the day of the interview/focus group to cover time and travel.

5. What are the possible disadvantages and risks of taking part?

There may be some issues discussed that may make you feel uncomfortable. For example, being diagnosed with HIV, transmitting HIV, and HIV partner notification (contacting sexual partners to advise them to test for HIV).

If you take part in a group discussion, this will be with a group of individuals who are also diagnosed with HIV or are also HIV negative gay or bisexual men. This means your current HIV status or sexuality may become known to them indirectly. Participants will be asked to keep the identities of the other participants in the group confidential. If you don't want to do a group discussion, you may be able to do a one-on-one interview instead.

6. What if something goes wrong?

If you are harmed by taking part in this research project, there are no special compensation arrangements. If you are harmed due to someone's negligence, then you may have grounds for legal action but you may have to pay for it. Regardless of this, if you wish to complain, or have any concerns about any aspect of the way you have been approached or treated during this study, please contact Larissa Mulka, investigator of the study (larissa.mulka@bsuh.nhs.uk). If you remain unhappy and wish to complain formally, you can do this by contacting Brighton Patient Advice and Liaison Service on (01273) 664511 or via email – pals@bsuh.nhs.uk.

7. What are the possible benefits of taking part?

There are no direct benefits to you personally. Your participation will help us determine how phylogenetics may be used in the future in a way that is acceptable to patients. It will help ensure concerns you raise regarding phylogenetics are addressed to prevent harm.

8. Will my taking part in this study be kept confidential?

Confidentiality and anonymity will be highly protected within the research team.

The only identifiable information taken from you will be your name and contact details, to help arrange the interview/focus group. These will be kept in a locked file, only the researcher has access to. All information taken from interviews or focus groups, including your questionnaire data will be collected under a study ID from which you will not be identifiable. Digital recordings will be typed up, then securely destroyed. The answers you give during the interview or discussion may be directly quoted in the study write up, but this will not be linked to any information which could reveal your identity.

Please note however, members of focus groups will share the same HIV status (see section 5).

9. What will happen to the results of the research study?

The results of the study will be presented at a medical conference and published in a medical journal. They may be used to guide further research. A lay summary will be produced which will be available at Brighton's HIV and Sexual Health clinic.

10. What will happen to my data?

Your data (name and contact details, and anonymised questionnaire and interview data) will be kept in separate password protected files on a secure university network for 5 years. This routine in medical studies. Any paper copies will be kept in a secure locked cupboard for 5 years, then all data will be safely destroyed.

10. Who is organising and funding the research?

This study is being funded by the British HIV Association and is Sponsored by Brighton & Sussex University Hospitals NHS Trust.

11. Contact for further information

You may ask questions about this study at any time. If you have any questions about the informed consent process or you require any further information relating to the study procedures or interviews, please contact Dr Larissa Mulka who is the researcher for this project. Email: Larissa.mulka@bsuh.nhs.uk

HIV/GUM Research Department, BSUH NHS Trust
The Elton John Centre
Sussex House
1 Abbey Road
BRIGHTON
BN2 1ES

Tel: 01273 523079

Fax: 01273 523080

Brighton and Sussex University Hospitals NHS Trust

CONSENT FORM

Title of project: A rapid ethical assessment of the use of phylogenetics for public health interventions in HIV.

Patient identification Number for this trial:

Names of investigator: Dr Larissa Mulka

Please initial box

1. I confirm that I have read and understand the information sheet dated 14.11.2017 for the above study and have had the opportunity to ask questions. ☐
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my medical care or legal rights being affected. ☐
3. I understand that my data will be digitally recorded then typed up. Things I say may be taken from the discussion and be directly quoted in the study write-up, but no identifiable information will be included in this. ☐
4. I agree to take part in the above study. ☐

Name of patient	Date	Signature

V 2.0, 14.11.17. IRAS ID 228210
Participant information and consent sheet, focus group discussion and interviews (patients)

Name of person receiving consent (if different from researcher)	Date	Signature
Researcher	Date	Signature

1 copy for participant, 1 copy for medical records (patient participants only); 1 copy for site file

Appendix F. Demographic questionnaires used for qualitative study (chapter 5)

Notes sheet

Age	
Sex	Male ☐ Female ☐ Trans ☐ Rather not say ☐
Ethnicity	
Sexuality	Straight/het' ☐ Gay/homosexual ☐ Bisexual ☐ Other ☐ Prefer not to say ☐
How often do you test for HIV?	At least every 6 months ☐ At least yearly ☐ Every 1 to 2 years ☐ Every 2 to 5 years ☐ Less than every 5 years ☐
Highest educational achievement (GCSE/A Levels/NVQ/Diploma etc)	

If there are any questions **you would rather not answer** please feel free to leave them blank.

Thank you.

Demographic Questions, HIV-negative participants. V 1.1, 04.09.2017

Notes sheet

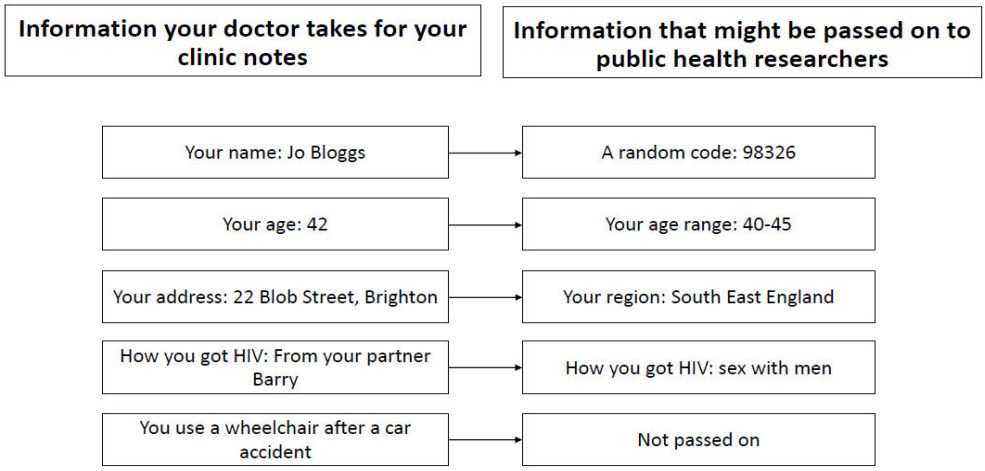
Age	
Sex	Male ☐ Female ☐ Trans ☐ Rather not say ☐
Ethnicity	
Sexuality	Straight/het' ☐ Gay/homosexual ☐ Bisexual ☐ Other ☐ Prefer not to say ☐
Risk factor for HIV	Sex with men ☐ Sex with women ☐ Injecting drugs ☐ Blood transfusion ☐ Mother to child transmission ☐ Prefer not to say ☐
Highest educational achievement (GCSE/A Levels/NVQ/Diploma etc)	

If there are any questions **you would rather not answer** please feel free to leave them blank.

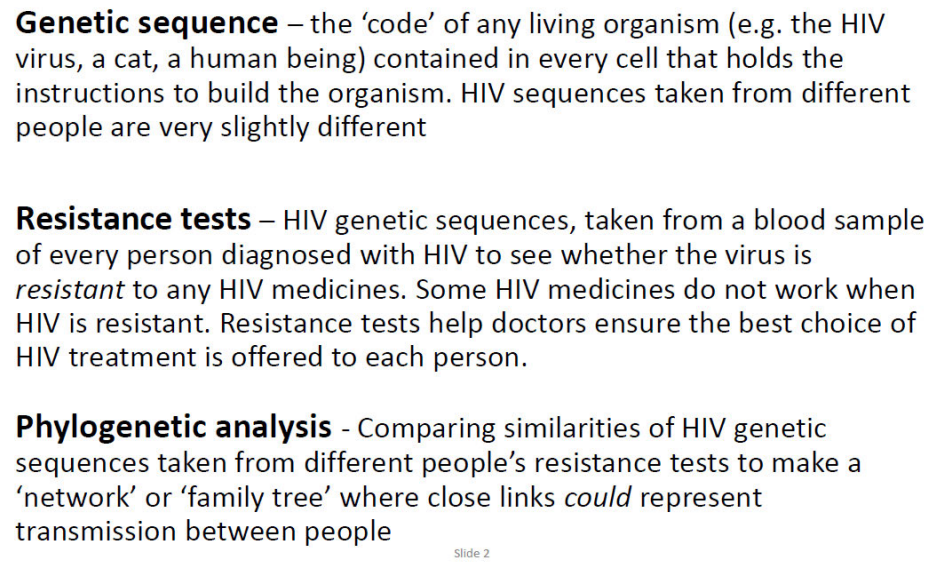
Thank you.

Appendix G. PowerPoint slides used to illustrate concepts and scenarios for qualitative study (chapter 5)

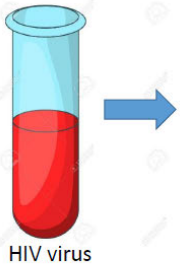
Topic 1: Awareness of use of personal data and anonymisation



Topic 2: Awareness and explanation of phylogenetics (3 slides)



HIV resistance test

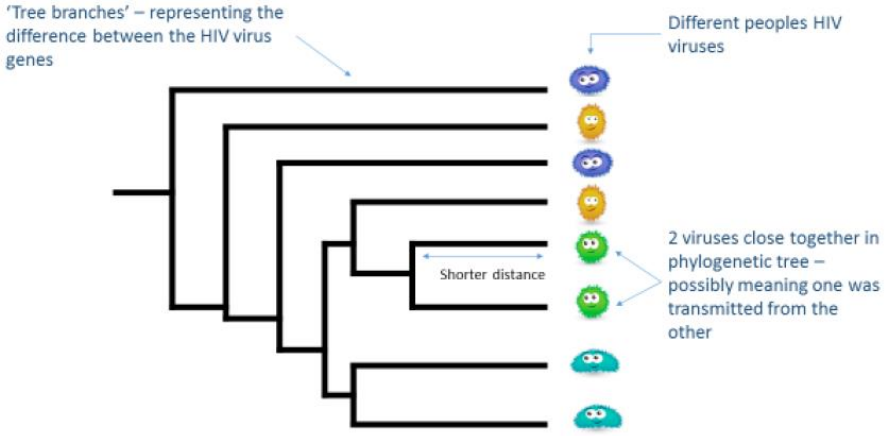


ACG GAG GGG TAT TTG TAG CCC CAA TTG GAG AAA GCC TAT TTT
CCC CAA TTG ACG GAG GGG TAT TTG TAG CCC CAA TTG GAG AAA
GCC TAT TTT CCC CAA TTG ACG GAG GGG TAT TTG TAG CCC CAA
TTG GAG AAA GCC TAT TTT CCC CAA TTG ACG GAG GGG TAT TTG
TAG CCC CAA TTG GAG AAA GCC TAT TTT CCC CAA TTG ACG GAG
GGG TAT TTG TAG CCC CAA TTG GAG AAA GCC TAT TTT CCC CAA
TTG ACG GAG GGG TAT TTG TAG CCC CAA TTG GAG AAA GCC TAT
TTT CCC CAA TTG ACG GAG GGG TAT TTG TAG CCC CAA TTG GAG
AAA GCC TAT TTT CCC CAA TTG TAT TTG TAG CCC CAA TTG GAG

Changes in the code here mean a certain drug won't work

Slide 3

Example of a phylogenetic 'tree':

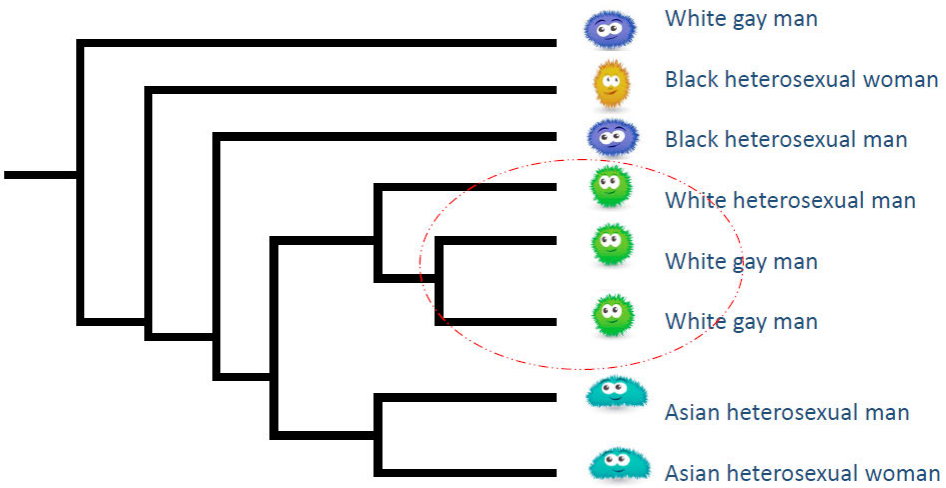


Slide 4

Topic 3: Anonymous phylogenetics (2 slides)

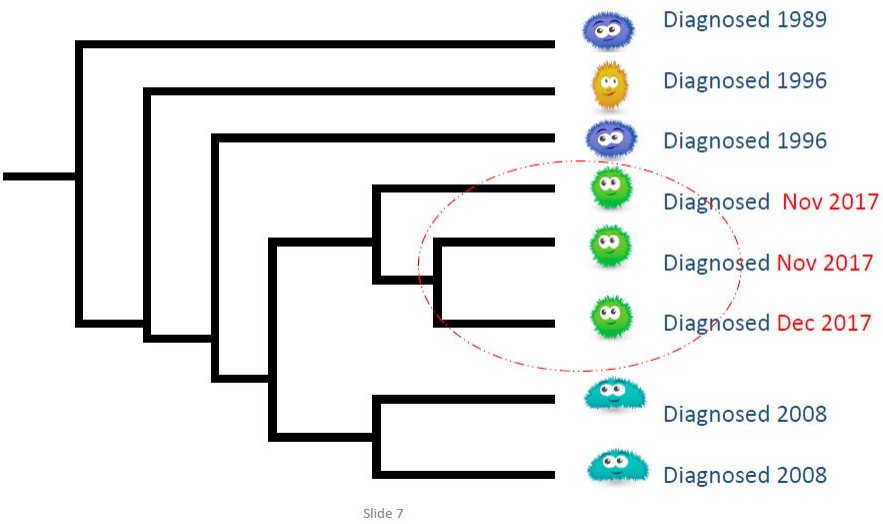
White, man, sex with men - ACG GAG GGG TAT TTG TAG CCC CAA TTG GAG AAA GCC TAT TTT
Black, woman, sex with men - ACG GAG GTG TAT TTG TAG CCC CAA TTG GAG AAA GCC TAT TTT
White, man, sex with women - ACG GAG GTG TAT TTG TAG CCC CAA TTG GAG AAA GCC TAT TTT
Black, man, sex with men - ACG GAG GGG TAT TTG TAG CGG CAA TTG GAG AAA GCC TAT TTT
White man, sex with women - ACG GAG GGG TAT TTG TAG CCC CAA TTG GAG AAA GCC TAT TTT
White, woman, sex with men - ACG GAG GGG TAT TTG TAG CCC CAA TTG GAG AAG GCC TAT TTT
White, man, sex with men - ACG GAG GGG TAT TTG TAG CGC CAA TTG GAG AAA GCC TAT TTT
White, man, sex with men - ACG GAG AAC TAT TTG TAG CCC CAA TTG GAG AAA GCC TAT TTT
Black, man, sex with women - ACG GAG GGG TAT TTG TAG CCC CAA TTG GAG AAA GCC TAT TTT
Asian, man, sex with men - ACG GAG GGG TAT TTG TAG CCC CAA TTG GAG AAA GCC TAT TTA
Asian, woman, sex with men - ACG GAG CCC TAT TTG TAG TTT CAA TTG GAG AAA GCC TAT TTT
Black, woman, sex with men - ACG GAG GGG TAT TTG TAG CCC CAA TTG GAG AAA GCC TAT TTT
White, woman, sex with men - ACG GAG GGG TAT TTG TAG GGC CAA TTG GAG AAA GCC TAT TTT
White, man, sex with men - ACG GAG GGG TAT TTG TAG CCC CAA TTG GAG AAA GCC TAT TTT

Slide 5

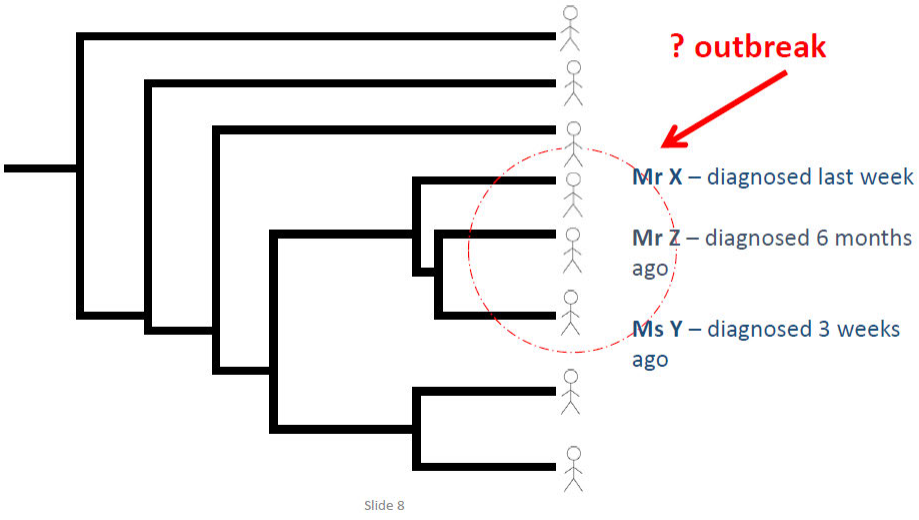


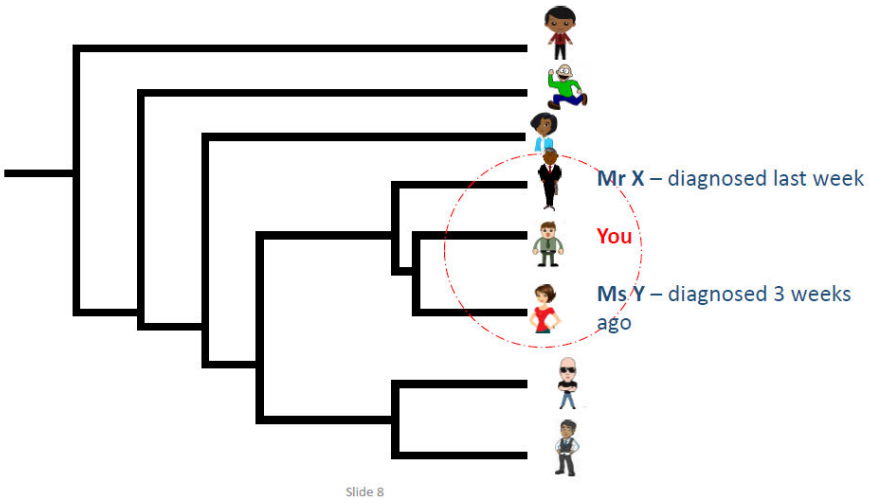
Slide 6

Topic 4: Anonymous phylogenetics in real time



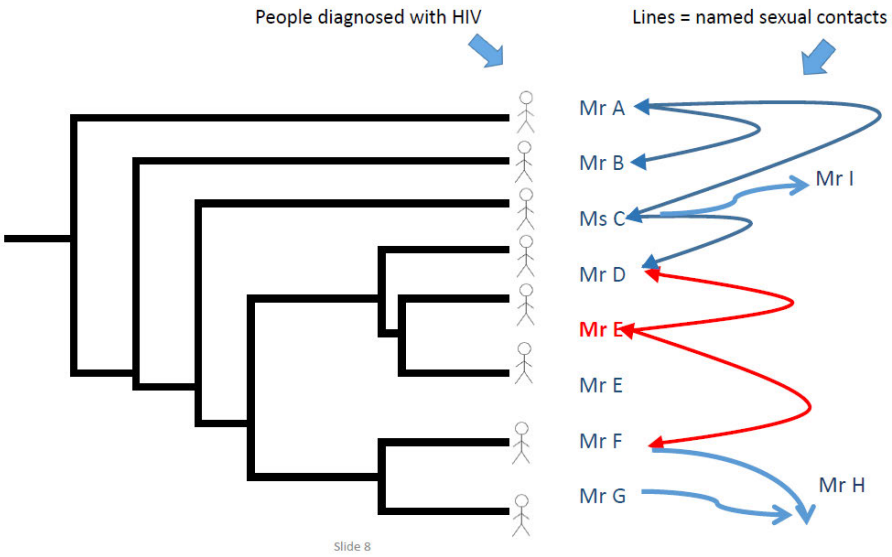
Topic 5: Real time phylogenetically led targeted interventions (2 slides)

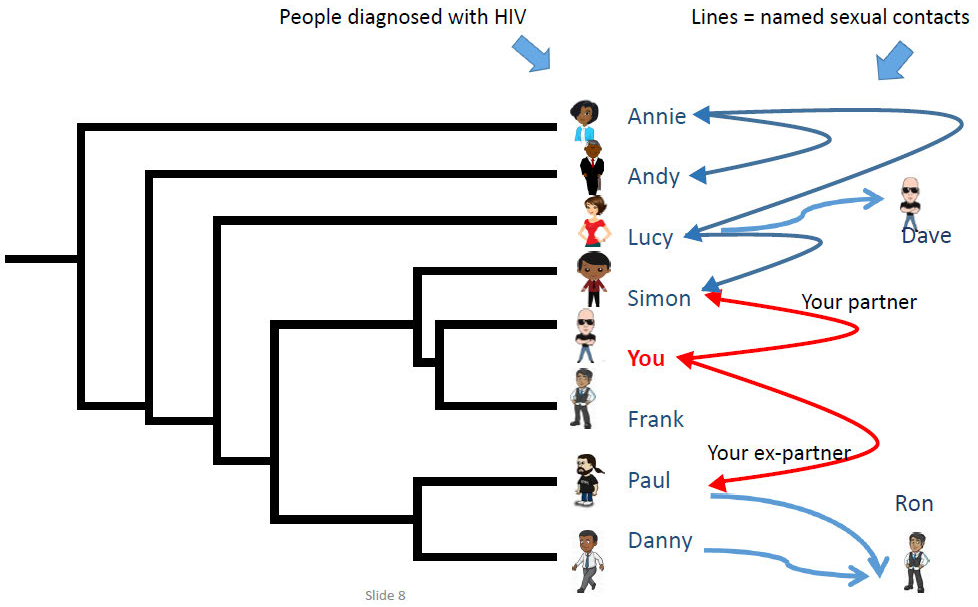




Topic 6: Advising a friend regarding HIV testing in the context of real time phylogenetically led interventions (no slides)

Topic 7: Linking partner notification data to phylogenetic networks (2 slides)





Topic 8: Next generation sequencing for phylogenetics (no slides)

Topic 9: Consent and public engagement (no slides)

Topic 10: Criminalisation (no slides)

Topic 11: Explanation of the law surrounding HIV transmission and debrief

HIV and the Law

It is *possible* to prosecute someone for transmitting HIV **intentionally** (on purpose) or **recklessly** (without taking care not to) if all of the following occur:

They know they have HIV
AND
They know HOW it is transmitted (passed on from person to person)
AND
They DO NOT TELL the person they might have transmitted HIV to that they are infected with it
AND
They do not take action to stop transmission happening (e.g. use condoms)
AND
HIV is actually passed on

Appendix H. Post discussion information sheet provided to participants in qualitative study (chapter 5)

Information summary – phylogenetics in HIV

Resistance tests

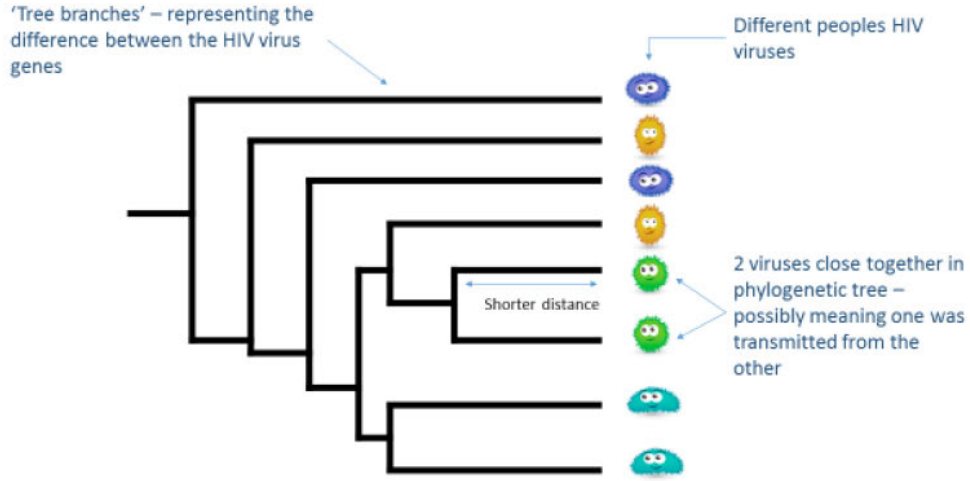
Resistance tests are HIV virus **genetic sequences** (the 'code' contained within HIV that holds the unique instructions to build the virus), taken from a blood sample of every person diagnosed with HIV to see whether the virus is resistant to any anti-HIV medicines (i.e. they won't work) to help guide treatment options.

These sequences are routinely **anonymised** (taking away any data that could link it to an individual) and stored in a national database with minimal linked data such as country of birth and risk factor for HIV, where they can be used for phylogenetic analysis.

Phylogenetic analysis

Phylogenetic analysis is comparing similarities of different peoples HIV genetic sequences to make a 'network' or 'family tree' where close links *could* represent transmission events between viruses.

Example of a phylogenetic tree:



1
V 1.1 04.09.2017 L Mulka

What can phylogenetic analysis tell us?

Phylogenetic analysis of sequences from resistance tests can tell us how similar different people's sequences are to give us an idea of which viruses are within the same network and may represent transmissions within the group, providing information to help with HIV prevention programmes. Standard phylogenetic analysis cannot tell us 'who-transmitted-to-who' for several reasons. Firstly, the sequences used are anonymised, so any viruses in the tree cannot be linked to the person it was taken from. Also, as we can't tell if there are other people missing from the phylogenetic tree which may have transmitted to one/both of the similar sequences, and from the tree, we can't see which way the virus was passed on (i.e. we can't tell if person A transmitted to person B, or Person B to person A). It gives us an overall idea of what is happening in the population.

HIV transmission and the Law

Though uncommon, it is *possible* to prosecute someone for transmitting HIV intentionally or recklessly if:

1. They **know** they have HIV

   AND

2. They **know how** it is transmitted (passed on from person to person)

   AND

3. They **do not tell** the person they might have transmitted HIV to that they are infected with it before they pass it on

   AND

4. They **do not take action to reduce the risk of transmission** happening (e.g. use condoms)

   AND

5. HIV is actually transmitted

Prosecutions are **very rare**, and are often only successful because the accused pleads guilty. The vast majority of accusations do not make it to court. A lot of information is needed as evidence including all previous partners of the accused and complainant, and it is very difficult to prove beyond reasonable doubt that one person infected another. Further information may be found through the Terrence Higgins Trust (http://www.tht.org.uk/myhiv/telling-people/law)

How can phylogenetics be used in court?

Phylogenetic analysis is used in criminal transmission cases. Usually in these circumstances, a new blood sample is taken from the accused and accuser for sequencing and an independent phylogenetic analysis is performed to see whether transmission may have occurred. These analyses can only say the two viruses are similar and CANNOT determine with certainty 'who-infected-who' as again, we can't prove that another person that wasn't included in the analysis transmitted to one or both of the people involved in the case, and can't tell the direction of transmission. We can however, say with certainty in some cases that transmission between two people did NOT happen, and this is how phylogenetics in court cases is most useful.

Could phylogenetics from my old resistance test be used to find out who my virus is linked to in the future?

The study you were involved with is trying to find out your thoughts on potential new ways of using resistance test sequences. This is based on work done in Canada, where standard phylogenetic analysis was performed in 'real time' (i.e. whenever anyone was diagnosed with HIV, their resistance test was taken and added to the 'tree' as soon as it was processed to see which groups were growing, as it happened). Where they suspected lots of transmission was occurring in a particular group, the individuals to whom the sequences belonged were identified (de-anonymised) and doctors approached them to make sure they were taking their medicine, and perform contact tracing with the individuals permission (find out who they had recently had sex with to ensure they tested for HIV) to try and reduce further spread, which appeared to have worked in one reported study. The investigators couldn't tell who infected who from the phylogenetic information, and people within each group did not have their identity revealed to others. In addition, the data used could not be used in court.

This use of phylogenetics IS NOT being used in the UK, and we feel it is very important that we have input from people living with HIV, and those who may test for HIV to see what their concerns might be and whether people testing for HIV would be comfortable with this process, before considering a similar strategy.

If a similar intervention was to be performed in the UK in the future, this would be done with guidance from studies including the one you have taken part in, and safeguards would be applied to make sure this was done in a way that was acceptable to patients and would not cause any harm to any individuals.

If you still have any questions regarding any of the information on this sheet, or relating to any aspect of the study, please contact the investigator, Larissa Mulka.
Larissa.Mulka@bsuh.nhs.uk