From the Division of Clinical Microbiology
DEPARTMENT OF LABORATORY MEDICINE
Karolinska Institutet, Stockholm, Sweden

MUTATIONS AND MIGRATION: MOLECULAR AND EPIDEMIOLOGICAL ASPECTS OF HIV-1 IN SWEDEN

Emmi Andersson

Karolinska Institutet

Stockholm 2019

Published by Karolinska Institutet.
Printed by Eprint AB 2019

ISBN 978-91-7831-541-3

MUTATIONS AND MIGRATION: MOLECULAR AND EPIDEMIOLOGICAL ASPECTS OF HIV-1 IN SWEDEN

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Emmi Andersson

Principal Supervisor:
Professor Anders Sönnerborg, MD PhD
Karolinska Institutet
Department of Laboratory Medicine
Division of Clinical Microbiology

Co-supervisors:
Professor Jan Albert, MD PhD
Karolinska Institutet
Department of Microbiology, Tumor and Cell Biology

Dr Johanna Brännström, MD PhD
Karolinska Institutet
Department of Medicine Huddinge
Unit of Infectious Diseases

Opponent:
Professor Karin J. Metzner
University Hospital Zürich
Department of Infectious Diseases and Hospital Epidemiology

Examination Board:
Professor Mika Salminen, PhD
National Institute for Health and Welfare, Finland
Department for Health Security

Docent Carl Johan Treutiger, MD PhD
Karolinska Institutet
Center for Infectious Medicine

Docent Hong Fang, PhD
Karolinska Institutet
Department of Laboratory Medicine
Division of Clinical Microbiology

Let the dataset change your mindset.

Hans Rosling

ABSTRACT

The Swedish HIV-1 epidemic is characterized by a wide variation in subtypes due to migration and travel. The aims of this thesis were to investigate molecular methods for estimating time to diagnosis and epidemiological tools for estimating the proportion of undiagnosed individuals with HIV-1 in Sweden, as well as to update the knowledge on transmitted and pre-treatment drug resistance (TDR and PDR) among persons living with HIV-1 in Sweden.

In **Paper I** we found that viral diversity estimated by the proportion of mixed base calls in *pol* gene sequences is an indicator of whether a HIV-1 infection is recent (≤1 year) or chronic. A cutoff of 0.47% mixed base calls identified recent infections with a sensitivity of 89% and a specificity of 75%.

In **Paper II** we studied TDR in newly diagnosed individuals (n=1,713, 71% coverage) in 2010-2016 in Sweden. The prevalence of TDR was 7.1% (95% CI 5.8-8.3%) and resistance to non-nucleoside reverse transcriptase inhibitors (NNRTI) increased significantly from 1.5% in 2010 to 6.2% in 2016 and was related to infection/origin in sub-Saharan Africa. A transmission cluster in men having sex with men with the M41L mutation was traced back to the 1990´s.

In **Paper III** we evaluated two mathematical models for estimation of the proportion of undiagnosed individuals with HIV-1 in Sweden, SSOPHIE and the ECDC Modelling tool. The model estimates for 2013 were 26% (n=2,100; 90% plausibility range 900-5,000) and 21% (n=2,013; 95% CI 1,831-2,189), respectively. The wide confidence bounds and problems to account for migration limited the value of the results. We concluded that improvements of models and better data on migration are needed.

In **Paper IV** we investigated PDR in individuals diagnosed in Sweden 2017-2019 (n=224) using an *in-house* high-throughput sequencing (HTS) assay with a detection limit of 1% for minority viral variants. HTS was successful in 87% of samples and failure was associated with low viral loads. Drug resistance mutations (DRMs) in protease/reverse transcriptase were detected in 49% of patients and in 18% of patients the DRMs were at levels ≥20% of the viral population. DRMs above 20% were mainly towards NNRTIs and correlated with previous ART exposure (p=0.001) and origin in Asia (p=0.007). Integrase strand transfer inhibitor (INSTI) DRM were rare and only found in levels <20% of the total viral population.

In conclusion, we have confirmed that viral diversity is a useful biomarker of time to diagnosis in HIV-infection and have identified important complications with estimating the proportion of undiagnosed HIV-1 infections in settings with high levels of migration. We have shown that the global increase in NNRTI-resistance is mirrored among HIV-positive migrants in Sweden and we have contributed to the knowledge about the utility of HTS in drug resistance testing of HIV-1.

POPULÄRVETENSKAPLIG SAMMANFATTNING

Syftet med min avhandling är att utveckla och studera metoder för att bättre kunna avgöra hur länge en individ varit smittad med humant immunbristvirus typ 1 (HIV-1) och hur många individer med odiagnosticerad HIV-1 infektion som vistas i Sverige. Detta sker genom undersökning av virusets arvsmassa och utvärdering av matematiska modeller. Vi ville också undersöka hur vanligt det är med överförd läkemedelsresistens och resistens innan behandling i Sverige. För att studera resistens i detalj har vi använt nya avancerade metoder för djupsekvensering som möjliggör fynd av även små mängder resistent virus.

HIV-1 har givit upphov till en världsomspännande epidemi och idag lever uppskattningsvis 37,9 miljoner människor med infektionen varav de flesta i Afrika. Viruset angriper immunförsvarets T-hjälparceller och utan behandling utvecklas i de flesta fall AIDS (förvärvat immunbristsyndrom) inom ett decennium. Eftersom HIV-1 under lång tid inte ger tydliga symtom kan en infektion förbli oupptäckt länge. Idag finns effektiv behandling med kombinationer av läkemedel som angriper virusets förmåga att föröka sig i kroppen, vilket kan eliminera smittsamhet och ger möjlighet till ett långt och friskt liv. Viruset har dock en inneboende förmåga att snabbt förändra sig (mutera) och om resistens mot läkemedel uppstår kan behandlingen bli ineffektiv.

Under 1980-talets första hälft var det framförallt män som har sex med män (MSM) och individer med intravenöst missbruk som infekterades med HIV-1 i Sverige. Migration och resor har förändrat bilden och andelen HIV-1 diagnoser hos individer födda i ett annat land och personer som blivit infekterade heterosexuellt har ökat kraftigt. Det senaste decenniet har cirka 450 individer per år diagnosticerats med HIV-1 och knappt 8000 individer lever idag med känd HIV-1 i Sverige. Tre fjärdedelar av nya HIV-1 diagnoser 2018 ställdes hos individer födda i ett annat land och andelen HIV-1 diagnoser hos MSM med utländskt ursprung ökar. I Sverige finns nu en stor variation av genetiska varianter (subtyper) av HIV-1, vilket speglar den svenska HIV-1 epidemins globala kopplingar.

De viktigaste slutsatserna från avhandlingen är:

- Information om virusets arvsmassa från rutintestning för resistens ger ledtrådar till hur länge en individ varit smittad.
- Uppskattningen av antalet odiagnosticerade individer med HIV-1 i Sverige är osäker och det behövs både utveckling av analytiska metoder och säkrare data om migration.
- Resistens mot NNRTI-läkemedel är hög hos HIV-positiva från Afrika söder om Sahara men finns även hos andra grupper vid HIV-diagnos i Sverige. Tidigare HIV-behandling och ursprung i Asien ökar risken för att bära på resistent virus vid HIV-diagnos i Sverige.

LIST OF SCIENTIFIC PAPERS

I. Andersson E, Shao W, Bontell I, Cham F, Cuong DD, Wondwossen A, Morris L, Hunt G, Sönnerborg A, Bertagnolio S, Maldarelli F, Jordan MR. **Evaluation of sequence ambiguities of the HIV-1 pol gene as a method to identify recent HIV-1 infection in transmitted drug resistance surveys.** *Infect Genet Evol. 2013 Aug;18:125-31. doi: 10.1016/j.meegid.2013.03.050.*

II. Andersson E, Nordquist A, Esbjörnsson J, Flamholc L, Gisslén M, Hejdeman B, Marrone G, Norrgren H, Svedhem V, Wendahl S, Albert J, Sönnerborg A. **Increase in transmitted drug resistance in migrants from sub-Saharan Africa diagnosed with HIV-1 in Sweden.** *AIDS. 2018 Apr 24;32(7):877-884. doi: 10.1097/QAD.0000000000001763.*

III. Andersson E, Nakagawa F, Van Sighem A, Axelsson M, Phillips AN, Sönnerborg A, Albert J. **Challenges in modelling the proportion of undiagnosed HIV infections in Sweden.** *Euro Surveill. 2019 Apr;24(14). doi: 10.2807/1560-7917.ES.2019.24.14.1800203.*

IV. Andersson E, Ambikan A, Brännström J, Aralaguppe S, Yilmaz A, Albert J, Neogi U, Sönnerborg A. **High-throughput sequencing for in-depth analysis of pre-treatment HIV-1 drug resistance in Sweden.** *Manuscript.*

ADDITIONAL RELEVANT PUBLICATIONS

I. Van de Laar MJ, Bosman A, Pharris A, Andersson E, Assoumou L, Ay E, Bannert N, Bartmeyer B, Brady M, Chaix ML, Descamps D, Dauwe K, Fonager J, Hauser A, Lunar M, Mezei M, Neary M, Poljak M, van Sighem A, Verhofstede C, Amato-Gauci AJ, Broberg EK. **Piloting a surveillance system for HIV drug resistance in the European Union.** *Euro Surveill. 2019 May;24(19). doi: 10.2807/1560-7917.ES.2019.24.19.1800390.*

CONTENTS

LIST OF ABBREVIATIONS

AIDS	Acquired immunodeficiency syndrome
ART	Antiretroviral therapy
AZT	Zidovudine
BIC	Bictegravir
bp	Base pair
CRF	Circulating recombinant form
DBS	Dried blood spot
DRM	Drug resistance mutation
DTG	Dolutegravir
d4t	Stavudine
ECDC	European Centre for Disease Prevention and Control
EFV	Efavirenz
EVG	Elvitegravir
gp	Glycoprotein
FTC	Emtricitabine
GRT	Genotypic resistance testing
HIV-1	Human immunodeficiency virus type 1
HIVDR	HIV drug resistance
HTS	High-throughput sequencing
IC_{50}	Median inhibitory concentration
IN	Integrase
INSTI	Integrase strand transfer inhibitor
LAg	Limited antigen assay
LMIC	Low- and middle-income country
MSM	Men having sex with men
NFLG	Near full-length genome sequencing
NGS	Next-generation sequencing
NRTI	Nucleoside reverse transcriptase inhibitor
NNRTI	Non-nucleoside reverse transcriptase inhibitor
NVP	Nevirapine

PDR	Pre-treatment drug resistance
PEP	Post-exposure prophylaxis
PHI	Primary HIV infection
PI	Protease inhibitor
PLHIV	People living with HIV
PMTCT	Prevention of mother-to-child transmission
PR	Plausibility range
PR	Protease
PrEP	Pre-exposure prophylaxis
PWID	People with intravenous drug use
RAL	Raltegravir
RT	Reverse transcriptase
SSA	Sub-Saharan Africa
SDRM	Surveillance drug resistance mutation
TAM	Thymidine analogue mutation
TDF	Tenofovir
TDR	Transmitted drug resistance
UNAIDS	Joint United Nations Programme on HIV and AIDS
URF	Unique recombinant form
USSR	Union of Soviet Socialist Republics
VL	Viral load
WHO	World Health Organization
3TC	Lamivudine

1 INTRODUCTION

Thirty-six years after the discovery of HIV-1 in 1983 and over 20 years after the introduction of efficient antiretroviral therapy (ART), the virus is still a global health challenge. Undiagnosed infections, lack of an efficient vaccine, and emerging antiretroviral resistance (HIVDR) are major obstacles to provide effective treatment for everyone in need and to stop the transmission of HIV-1 worldwide.

An estimated 37.9 million people were living with HIV-1 and 23.3 million were on ART at the end of 2018. The most affected region is eastern and southern Africa with 20.6 million people living with HIV-1, whereof 13.8 million on treatment. [1] However, the virus is spreading rapidly also in countries like Russia and China. [2] Modern ART has transformed HIV-1 from a deadly infection into a treatable chronic condition with limited or even eliminated risk of transmission [3-5], but still no cure or protective vaccine is available. In 2018, 770,000 people died from HIV-1 infection, a decline with 33% since 2010. [1]

Sweden is a low prevalence and low incidence country for HIV-1 but mirrors the world in that migrant populations from high endemic countries are well-represented in the Swedish national HIV-1 positive population. By studying HIV-1 in Sweden, we thereby contribute to knowledge of HIV-1 both nationally and globally.

1.1 HIV-1 VIROLOGY

1.1.1 The virus and its origin

HIV-1 belongs to the genus *lentiviridae* within the *retroviridae* family. The single-stranded positive sense RNA-genome has a length of approximately 10,000 base pairs (bp) and contains nine genes in three reading frames. The three major genes are: *gag*, that encodes the major structural proteins (matrix protein p17 and capsid protein p24), *pol* that encodes the enzymes reverse transcriptase (RT), integrase (IN) and protease (PR), and *env* that encodes envelope glycoproteins (gp41 and gp120). The genome also contains six genes that encode proteins with regulatory function; *tat, rev, vif, vpr, vpu and nef*. Both 3´and 5´ends of the RNA genome contain LTRs (long terminal repeats) that are important for proviral insertion in the host genome after reverse transcription as well as for replication, transcription and translation. [6]

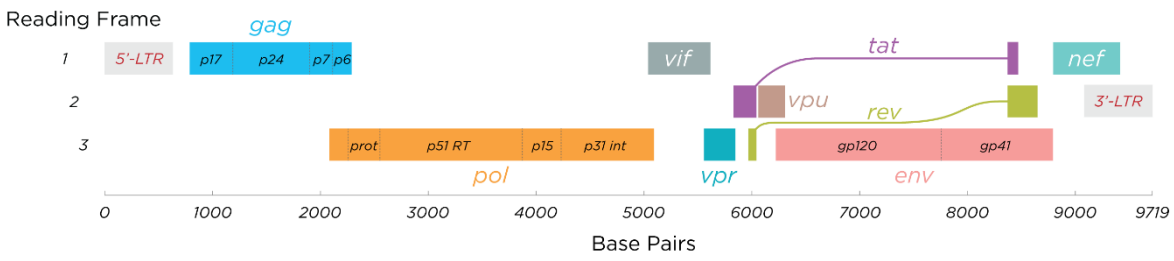

Fig 1. The HIV-1 genome. *Wikimedia Commons. Thomas Splettstoesser* (www.scistyle.com)

The viral particle contains two copies of the genome within a conical nucleocapsid. The nucleocapsid is surrounded by a host cell derived envelope with viral proteins gp120 and gp41. The virion also contains the viral enzymes RT, PR and IN [7]

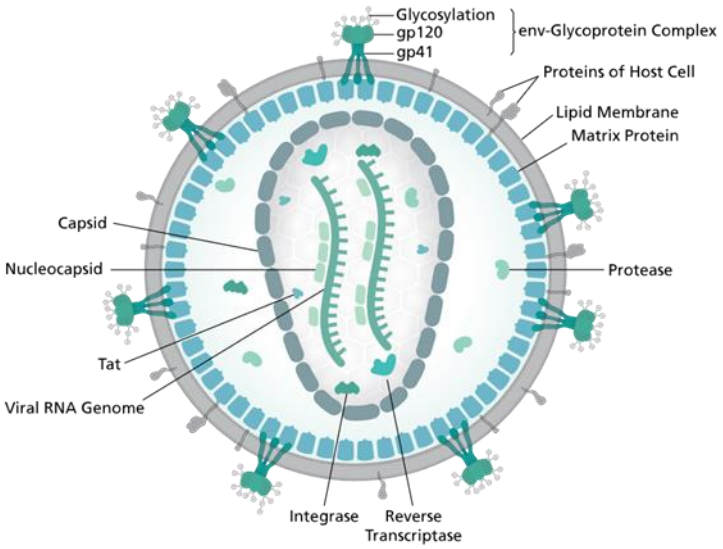

Fig 2. The HIV-1 virion. *Wikimedia Commons. Thomas Splettstoesser* (www.scistyle.com)

HIV-1 originates from the simian virus SIV that is found in chimpanzees and gorillas in Cameroon. Multiple cross-species introductions of SIV in humans have resulted in four known groups related to different zoonotic events (M, N, O and P). [8] HIV-1 group M (M for "major") is the pandemic branch of HIV-1 that represents most infections in the world and is estimated to originate from transmission of SIV to man in Central Africa in the early 1900s. [9, 10] The pandemic spread of group M within Africa is not only related to the virus itself, but to sociodemographic factors during the early 20th century such as migration [11, 12], unclean practises in medicine [13], increased prevalence of genital ulcer disease [14] and exploitation of the region by European colonizers establishing transportation networks and inferring social changes. [12] From the local epidemic in Congo in the 1960´s, the virus was brought to Haiti with returning workers/soldiers, and then to North America probably in the late 1960´s. It is believed to have circulated in the USA for around 12 years, before surfacing as a clinical condition when the first cases of AIDS were described in 1981. [15]

The closely related HIV-2 has resulted from zoonotic transmission of a distinct variant of SIV from sooty mangabey monkeys. HIV-2 is less pathogenetic and transmissible and has mostly spread in West Africa. [8]

1.1.2 The HIV replication cycle

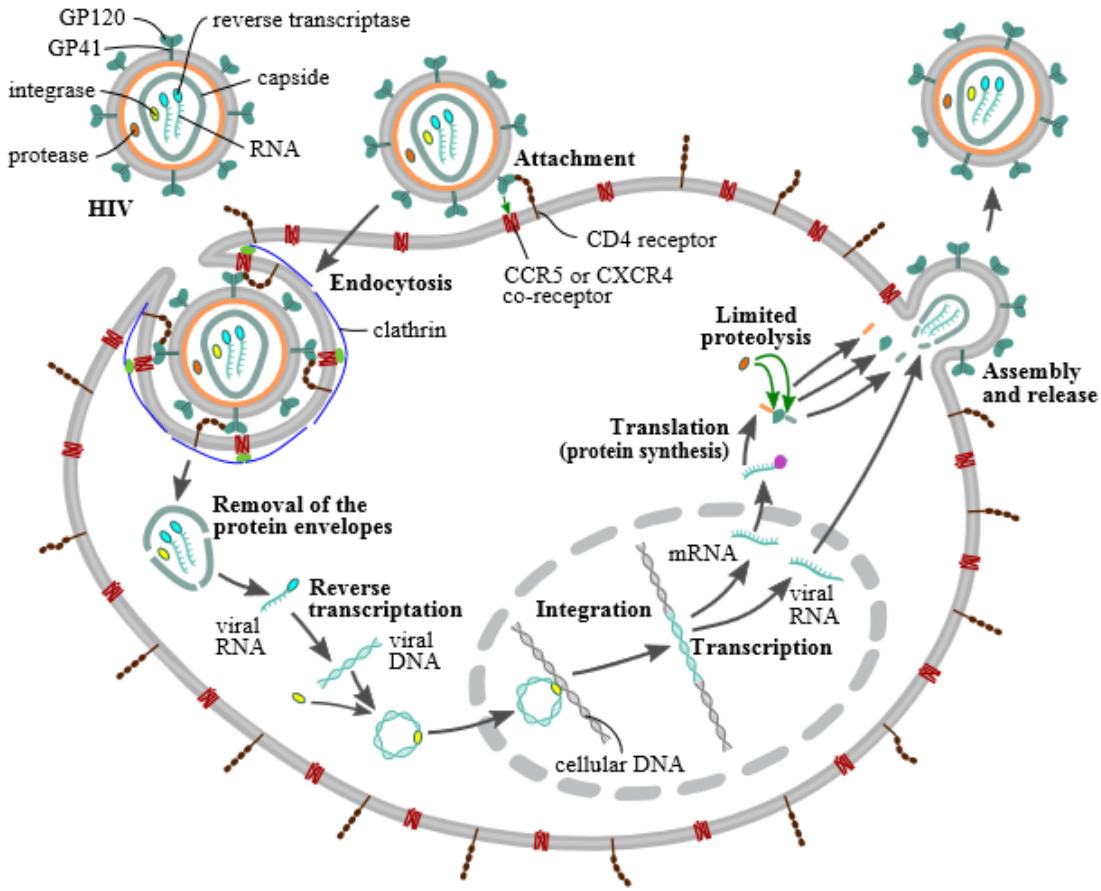

Fig 3. The replication cycle of HIV-1. *Wikimedia Commons. Jmarchn.*

HIV-1 infects cells bearing the CD4-receptor, the main receptor of the virus. This makes $CD4^+$ T-cells the primary target of infection, but macrophages, dendritic cells, microglia and astrocytes can also be infected. [16]

Glycoprotein 120 (gp120) on the virion surface binds with high affinity to the CD4-receptor, and conformation changes in the gp120 then allows for interaction with one of the chemokine co-receptors CCR5 or CXCR4. Structural rearrangements of gp41 then lead to fusion of the viral envelope with the cell membrane and subsequent HIV entry. [17, 18]

Transmitted viral variants are almost exclusively CCR5-tropic and in early infection usage of this coreceptor dominates. In advanced disease, however, a switch to CXCR4 use may occur, but this is not obligate for development of AIDS. [19]

When HIV-1 has entered the target cell, the RT enzyme creates a double stranded DNA-copy of the viral genome, that is subsequently imported into the nucleus and integrated into the host genome by the integrase enzyme.

The integrated HIV-1 provirus is the template for the viral RNA transcripts, that are produced by the cellular RNA polymerase II enzyme. New copies of the viral genome are produced as

well as several different species of mRNAs that are translated into viral proteins in the cytoplasm. The assembly of the genome and viral proteins at the plasma membrane leads to budding of viral particles from the host cell. The last step to render the released virions infectious is the action of the protease enzyme that cleaves Gag-Pol polyproteins into functional mature proteins by cleavage of peptide bonds. [6, 7]

The integration of proviral HIV-1 in the host genome builds a latent reservoir in inactivated and long-lived $CD4^+$ memory cells. Even though only a few percent of integrated HIV-1 proviral sequences are functional [16], these suffice to reignite the infection if treatment is discontinued. Therefore, ART must be life-long. To eradicate or inactivate this functional viral reservoir is one of the main focuses of HIV-1 cure research. [16, 20]

1.1.3 Mechanisms of viral diversity

Genetic diversity, both within-host and between-host, is a key characteristic of HIV-1. The molecular basis of viral diversity is mutations and recombinations of viral genomes, followed by selection of the most successful variants in a certain environment. The mutation rate in HIV-1 is extremely high, about 2×10^{-5} mutations per base per replication cycle in cell culture experiments, confirmed by *in vivo* analyses of neutral and deleterious mutations in whole genome deep sequencing data. [21] Mutations are caused by errors introduced by the RT and, to lesser degree, the human RNA polymerase II as well as by host cellular factors such as APOBEC-proteins that exert an antiviral effect by hypermutating viral sequences. [21, 22] The high mutation rate leads to production of many non-functional virions, but occasionally to successful mutations that improve virus fitness, and therefore are propagated through natural selection.

Recombination in HIV-1 can happen when two viral variants infect the same cell, and their RNA progeny is co-packaged into one virion. When such a viral particle, with two genetically distinct RNA molecules, infects another cell, the inherent template switching of the RT results in a chimeric DNA-provirus. Recombination adds to genetic variability and can move accumulated combinations of mutations between viral variants. [23] Recombination is important within the context of both drug resistance development and viral diversity on the global level.

1.1.4 Diversity between hosts - subtypes

HIV-1 group M has diversified into genetic subtypes A-D, F-H, and J-K since its transmission from chimpanzees. [10, 24] Subtypes are defined by phylogenetic clustering of viral sequences and divided into sub-subtypes (e.g. A1, A2 and A6 within subtype A). Genetic variation within a subtype is about 8-17% and between subtypes usually 17-35%. Recombinations between subtypes may occur in dually infected patients and recombinant variants detected in at least three non-linked persons are named circulating recombinant forms (CRFs), e.g. CRF01_AE. Mosaic recombinants with genome fragments from four or more subtypes are called complex (cpx) (e.g. CRF04_cpx). Currently there are 97 CRFs

listed on Los Alamos HIV sequence database´s webpage (www.hiv.lanl.gov, accessed August 20, 2019). Unique viral sequences are named unique recombinant forms (URFs) and are reclassified into CRFs if found in three or more epidemiologically unlinked individuals. [25, 26] Most subtyping has so far been based on partial *pol* sequences of approximately 1,000 bp, because such sequences are generated by routine HIV-1 genotypic resistance testing (GRT). The categorization of HIV-1 strains is challenged by the constant evolution of new variants and needs to be flexible. Today there is an increasing amount of near full-length genome (NFLG) data that will add information on the complexity of HIV-1 evolution, since analysis of larger parts of the genome detect inter-subtype recombinants and URFs with higher sensitivity. [27, 28]

The different subtypes have so far been highly associated with geographic areas, probably due to a combination of founder effects, regional migration, population growth and urbanization. [26] A recent systematic review of available data shows that the patterns of HIV-1 diversity globally are dynamic and that the proportion of recombinants is increasing over time. Subtype C is the most common subtype worldwide (46.6 %), followed by B (12.1%) and A (10.3%). [29] Subtype C was described for the first time 1986 in Ethiopian patients by Prof Sönnerborg´s research group [30] and is likely to have spread from Katanga in Congo. Central Africa has the most diverse collection of subtypes and recombinants, reflecting the origin of all HIV-1 variants in this region. In east and west Africa, the epidemic is dominated by subtype A, and CRF02_AG and G respectively. In southern Africa, Ethiopia and India subtype C is found almost exclusively. [29, 31] North America, the Caribbean, Latin America, western and central Europe and Australia are dominated by subtype B, but CRFs and URFs are increasing. In eastern Europe and central Asia subtype A is most common, but with co-circulation of B and CRFs. [29] In this region the sub-subtype A6 dominates. [32] In east- and southeast Asia, CRF01_AE is most common and in the Middle East and north Africa, subtype B is decreasing and CRFs, mainly CRF35_AD have increased to more than half of the infections. [29]

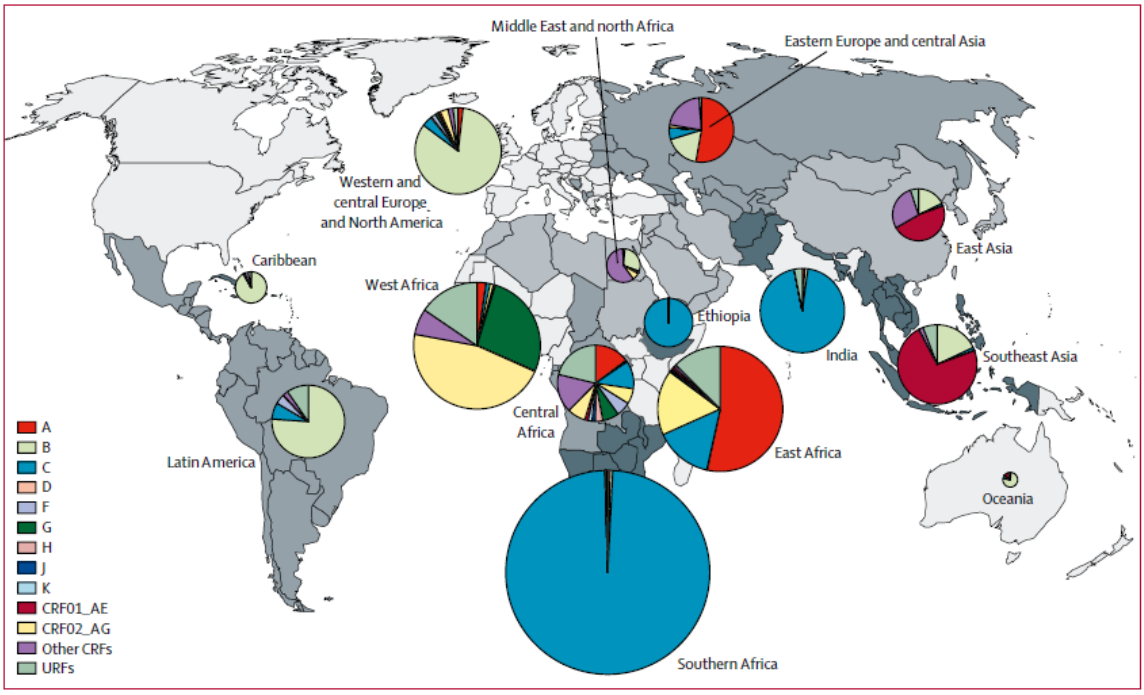

Fig 4. Worldwide distribution of HIV-1 subtypes 2010-2015. Reprinted with permission. [29]

Since the beginning of the HIV-1 epidemic, subtype B has dominated in western Europe. It is still the most common subtype, but introduction of other subtypes and CRFs through migration and travel have resulted in increased molecular heterogeneity and unique subtype compositions of the HIV epidemics in different countries in western and central Europe. In eastern Europe the epidemic did not gain momentum until the early 1990´s, but then an explosive and homogenous spread of subtype A (especially sub-subtype A6) was fuelled by transmission in people with intravenous drug use (PWID). In more recent years it has expanded into the heterosexual population. [33] In the Nordic countries (Sweden, Denmark and Finland) subtype B is declining and the proportion of CRFs is increasing. [27, 34]

1.1.5 Diversity within a host - viral quasispecies

HIV-1 exists in the infected individual as a collection of viral variants called a quasispecies, that is constantly reshaped by evolutionary selection. In the infected individual the viral variability enables it to evade the immune system, develop drug resistance and escape vaccine strategies [23, 35, 36] Transmission and primary HIV-1 infection is in most patients caused by a single viral variant, and transmucosal transmission is suggested as the bottleneck event. [37-40] Recent data on higher viral diversity in the female genital tract than in blood in early infection suggests that this bottleneck event happens when infections in the cervical mucosa become systemic in female heterosexual infection. [41, 42] Viral variants with high fitness are favoured in transmission. As the founder virus replicates high viral load (VL), often > 1 million RNA copies/ml, is usually reached during acute infection. [23, 37, 43]

Viral diversity in the blood then increases fast in a linear fashion, reaches a plateau and finally, when $CD4^+$ T-cell counts are failing in advanced disease, both divergence from the

founder strain and diversity may decrease. [44]. The founder sequence is the starting point for the viral evolution that explores the sequence space around it, and in advanced disease the viral populations tend to approach a global HIV-1 consensus. In a large replicating population, the main limitation of genetic diversity is the fitness cost of combined mutations. [45] As the dominant viral variant in plasma change due to changes in selective pressure, less favoured variants may still be present in minority viral variants. A special feature of retroviral quasispecies is that viral variants will be archived as proviral sequences and can be resurrected in a later stage of infection. [46, 47]

1.2 HIV-1 EPIDEMIOLOGY

1.2.1 The state of the HIV-1 pandemic

HIV-1 has spread successfully throughout the world through sexual transmission, transmission from mother to child, intravenous drug use, unclean medical practises and untested blood products. When the virus was identified in 1983, the pandemic was already established. Today, 37.9 million people worldwide are estimated to be living with HIV and thanks to massive efforts to improve access to treatment, 23.3 million of them are on ART. [1] The worldwide spread of the common viral subtypes has been retrospectively elucidated with phylogeography, using the phylogenetic relationships between viral sequences in a geographic framework. HIV-1 does not spread at random from high-endemic areas, but the spread is directed by geopolitical factors. [48, 49]

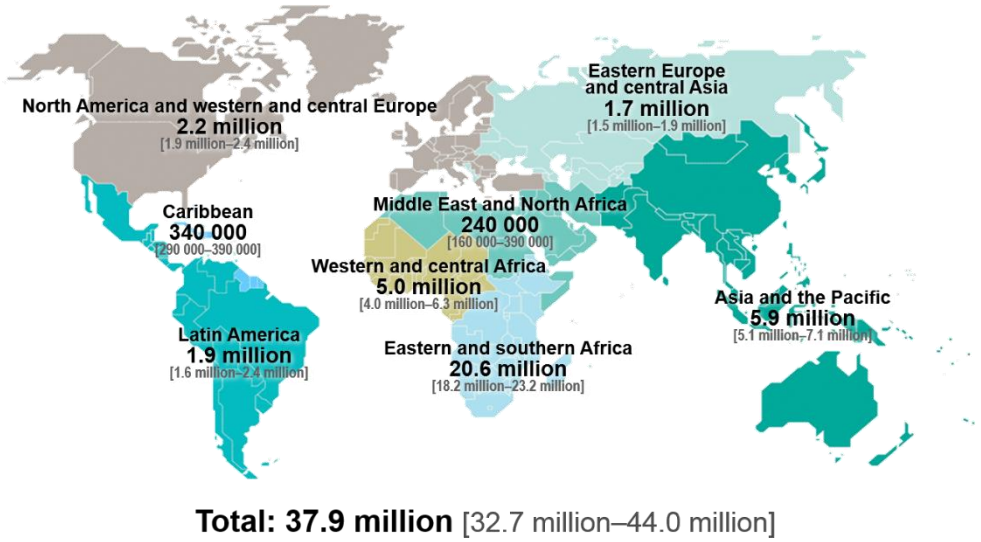

Fig 5. The global distribution of people living with HIV in 2018. Reprinted with permission. https://www.unaids.org/en/resources/documents/2019/core-epidemiology-slides

The nature of the epidemic differs between countries and populations. Key populations within people living with HIV (PLHIV) worldwide that are highlighted by the World Health Organization (WHO) include men having sex with men (MSM), people with intravenous drug use (PWID), sex workers, transgender people and people in prison. In 2018, an

estimated 54% of all new HIV infections affected individuals in these key populations or their sex partners. Individuals in these groups have a higher risk of HIV infection in all parts of the world but are also vulnerable and facing social and legal barriers to access diagnosis and care. In many countries with generalized HIV epidemics, these groups have not been prioritized. [50, 51]

The area that is most affected by HIV-1 is sub-Saharan Africa (SSA), in particular eastern and southern Africa, where the epidemic is mainly driven by heterosexual transmission and the estimated HIV-1 prevalence in adults of reproductive age (15-49 years) in 2018 was 7.0%. The prevalence differs in the region and is highest in South Africa (20.4%), Botswana (20.3%), Lesotho (23.6%) and Eswatini (27.3%). [52] The number of PLHIV is estimated to be highest in South Africa (7.7 million), Mozambique (2.2 million) and Nigeria (1.9 million). [52] Regional and national estimates of prevalence are only averages, and to understand the dynamics of HIV transmission breakdown of data in a location-population approach has been suggested. Sub-national diversities in prevalence and the importance of geographical high-prevalence hot spots for HIV transmission have recently been demonstrated and calls for more customized interventions to control HIV-1 in SSA. [2, 53, 54]

The general prevalence of HIV-1 in SSA is higher in women than in men. Women and girls are at higher risk for HIV infection because of gender inequalities, gender-based violence and in some countries policies and laws limit access to reproductive health services for minors. Efforts to eliminate mother-to-child transmission have been successful with 92% of pregnant women living with HIV in eastern and southern Africa receiving antiretrovirals during pregnancy, resulting in an average rate of transmission from mother-to-child under 10%. However, the use of single drug nevirapine in this setting has also resultated in a high rate of drug resistance among mothers and in the children in whom the prophylaxis has failed. [55, 56] Other remaining issues are the high number of new infections during pregnancy that go undiagnosed and untreated, and that diagnosis in infants is often delayed when HIV-1 RNA point-of-care testing is not available (68% of infants born to HIV-positive women were tested within 2 months from delivery). New HIV infections and AIDS related deaths decline at the highest pace in the world, as a result of the roll out of ART, but the epidemic remains massive. [1, 52]

The epidemic in North America, where the first cases of AIDS were diagnosed in the 1980´s, is closely linked to the epidemic in western and central Europe. HIV-1 was introduced in Europe through intravenous drug use and MSM contacts in or from USA and through heterosexual contacts from Central Africa. The countries with most non-European influx of subtype B were UK, France and Switzerland. [48, 57] The high-income countries in North America and western and central Europe have a high coverage of HIV-services and feasibility of individualized monitoring and treatment of PLHIV. MSM are still the group that is most affected by HIV-1 and 57% of new infections in 2017 occurred in this group. Remaining issues in the western world are a high rate of late diagnosis, 48% of individuals diagnosed with HIV-1 in the European Union and the European Economic area in 2016 had a

$CD4^+$ cell count below 350 cells/µl at diagnosis. There is still widespread stigma and discrimination within the health care systems in Europe, creating barriers for testing and treatment especially for sex workers, MSM and PWID. [2]

In the eastern European countries that were behind the iron curtain until the end of the 1980ies and in the former USSR, the introduction of HIV-1 was delayed due to geopolitical isolation. In 1994 an explosive spread of a subtype A clade (sub-subtype A6) with probable origin from central Africa was noted among PWID in Ukraine and subsequently in Russia, Belarus and other countries in the area. [32, 33, 58] The HIV-1 epidemic is now large and has grown by 30% since 2010 with a majority of PLHIV (70%) living in the Russian federation. In 2017 the estimated number of PLHIV in Russia was 1 million. Harm reduction programmes for PWID are insufficient and people who inject drugs and their sexual partners are the major risk group for new HIV-1 infections. Infections in MSM are on the rise but testing in this group represent a disproportionally small part of all tests, related to harsh attitudes and policies towards same sex relations. [2]

In Latin America, most new infections occur in MSM, but commercial sex is also an important transmission route. In some countries the HIV prevalence in MSM and transgender women is very high. [2] Latin America is identified as a "donor" of subtype B infections to both North America and Europe. [48]

In South East Asia the dominating HIV-1 subtype is the recombinant CRF01_AE, and phylogeographic studies show that it spread from central Africa to Thailand, probably through heterosexual contacts. A rapidly expanding epidemic was established in Thailand in the early 1990ies and lead to spread to other countries in Asia as well as to Europe and North America. Concomitantly subtype B was introduced into the region through persons from USA. Thailand is still an important hub for transmission of HIV-1 and its importance is explained by its popularity as travel destination in combination with a large sex work industry. [49, 59]

The response to the epidemic in the Asian countries is diverse. The estimated number of PLHIV is highest in India (2.1 million, 2017) [2], China (1.25 million) [60], Indonesia (640,000, 2018) and Thailand (480,000, 2018). [2] Several countries have had success in reducing the transmission related to sex work, but transmission among MSM is of increasing importance and rising. New infections and AIDS related deaths in the whole area are decreasing, but in Pakistan and the Philippines the epidemics are expanding and new infections in young people are on the rise. In both countries sex work is criminalized and parental consent is needed for HIV testing in individuals <18 years old. In Pakistan, same sex acts can lead to a death sentence. [2]

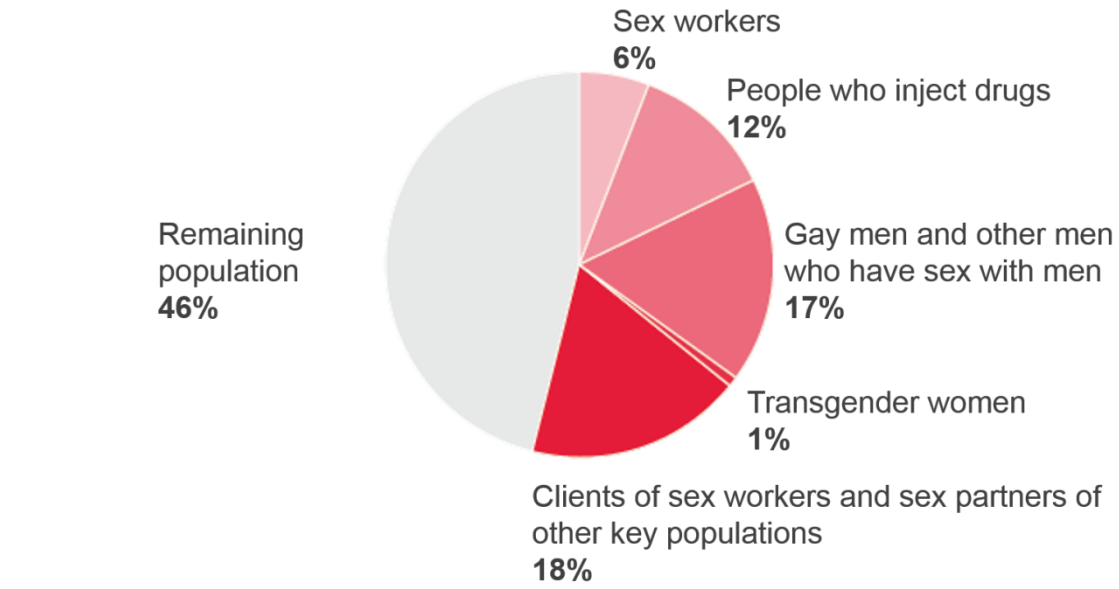

Fig 6. Distribution of new HIV infections 2018 globally by key population. Reprinted with permission. https://www.who.int/hiv/en/

1.2.2 The global 90-90-90 treatment target

In 2014 the Joint United Nations Programme on HIV and AIDS (UNAIDS) stated that advances in ART to treat and prevent HIV infection make it possible to end the AIDS epidemic by 2030. [61] To concretize an intermediate goal, the 90-90-90 treatment target to be reached by 2020 was formulated.

The target states that:

- "By 2020, 90% of all people living with HIV will know their HIV status."
- "By 2020, 90% of all people with diagnosed HIV infection will receive sustained antiretroviral therapy."
- "By 2020, 90% of all people receiving antiretroviral therapy will have viral suppression."

To summarize, these goals would equal at least 73% of all PLHIV being on effective ART by 2020. [62]

A "fourth 90" has been suggested since PLHIV with viral control still experience challenges posed by co-morbidities and HIV-related stigma. This target would "ensure that 90 % of people with viral suppression have good health-related quality of life". [63]

The UNAIDS "Fast-track targets to end the AIDS epidemic by 2030" also contain a target of 95-95-95 by 2030 and total elimination of HIV-related discrimination. [64]

The target objectives are modelled on "the HIV care continuum" or the "treatment cascade", with principles suggested already in 2005 to give an integrated description of HIV care

including diagnosis, linkage to care and retention in care and on ART as important aspects of successful treatment. This is an efficient tool to identify gaps in service delivery to PLHIV. [65, 66] WHO have recently issued practical guidelines on collection and management of cascade data. [67]

The first country that reported to have achieved the 90-90-90 goal was Sweden. [68] Other countries in Europe that have also achieved the target goals are Denmark, United Kingdom, Switzerland, Germany, The Netherlands, Austria and France. [69] The global figures today are 79-78-86 with a resulting 53% of all PLHIV reaching the goal of viral suppression. [1] National and regional differences are huge and both epidemiological conditions and health care structures differ between countries, so responses and actions need to be customized nationally. Many challenges are however the same, to start with, to obtain reliable measurements of all the target objectives. Most countries do not have full coverage of HIV incidence reporting or full records on PLHIV on treatment and mathematical modelling of cascade data is a necessity. In low income settings VL is not measured in all patients. [67]

1.2.3 HIV-1 in Sweden

The general prevalence of HIV-1 in Sweden is low (0.08%). (InfCare HIV Nov 20, 2018) Since 1983 reporting of first AIDS and when diagnostic tests became available, HIV, has been mandatory and almost 13,000 accumulated cases have been reported. (Public Health Agency of Sweden) During the last decade, an average of 450 new cases of HIV-1 per year has been registered. [70] All 7,953 individuals currently under HIV care (InfCareHIV Aug 21, 2019) are included in the national HIV database InfCareHIV, where epidemiological and clinical data is collected in real time. The database has full national coverage since 2009. [71]

The first case of HIV-1 in Sweden was identified retrospectively to have been in care in 1976 and the first AIDS case was identified in 1981 at Roslagstulls Hospital, Stockholm. Both individuals were heterosexually infected (Personal communication, Anders Sönnerborg). Retrospective analysis of HIV-1 antibodies has shown that the virus was introduced among MSM late 1979. [72] HIV-1 had an epidemic spread in Sweden in the beginning of the 1980´s, first in MSM and somewhat later in PWID. The spread mainly took place in small groups of individuals with high-risk behaviour. When the individuals in these groups, estimated to some hundred PWID and less than a thousand MSM, were already infected, the spread decreased even before active prevention measures were in place. The heterosexual spread within Sweden was limited but several AIDS cases were identified during the second half of the 1980´s and related to migration. [73]

In 2018, 481 cases of HIV-1 were reported in Sweden, including both entirely new diagnoses and migrants entering Swedish HIV-care with a previous diagnosis abroad. Sixty-four percent of the diagnoses were in men and the median age at diagnosis 38 years (range 1-77). In 76 cases HIV transmission within Sweden was reported, and 39 of those were in MSM. [70]

The characteristics of individuals with new HIV-1 diagnoses in Sweden have shifted significantly over the years and migration and travel are now important factors. In 2018, 75% of individuals reported with HIV were born abroad. Migrants from SSA with heterosexually acquired infections and migrants from Thailand are important groups and there is an increase in the proportion of migrant MSM amongst those diagnosed with HIV-1 in Sweden. [70] In the ongoing TIME-study on newly diagnosed HIV-infection in Sweden, 53% of 106 included MSM so far were of non-Swedish origin. (Personal communication, Emmi Andersson)

The most common countries for HIV transmission abroad in 2018 were Thailand and Eritrea. [70] However it is not unlikely that persons who officially origin from Eritrea have their origin in Ethiopia, as a result of misclassification. Interestingly, these countries are not reported to have exceptionally high HIV-1 prevalence (Eritrea 0.7%, Ethiopia 1.0%, Thailand 1.1%) [52] suggesting that migration is not a random event and that epidemiology of sub-populations as well as the migration process itself is important when considering HIV-risk. An increasing proportion of individuals diagnosed with HIV-1 in Sweden have already tested positive for HIV-1 abroad and many are on successful ART at arrival, but this information is not completely captured in surveillance data and the trend is based on clinical observations.

In individuals originating from high-endemic settings and diagnosed with HIV-1, transmission is often presumed to have taken place before migration to Sweden. However, a study using a $CD4^+$ trajectory model to estimate time of infection concluded that 10% of migrants considered to be infected before arrival most likely got their HIV-1 infection after arrival to Sweden, and that transmissions among migrants after arrival to Sweden might be underestimated. [74] A similar pattern has been shown in Europe for both migrants who have been heterosexually infected and MSM. [75, 76]

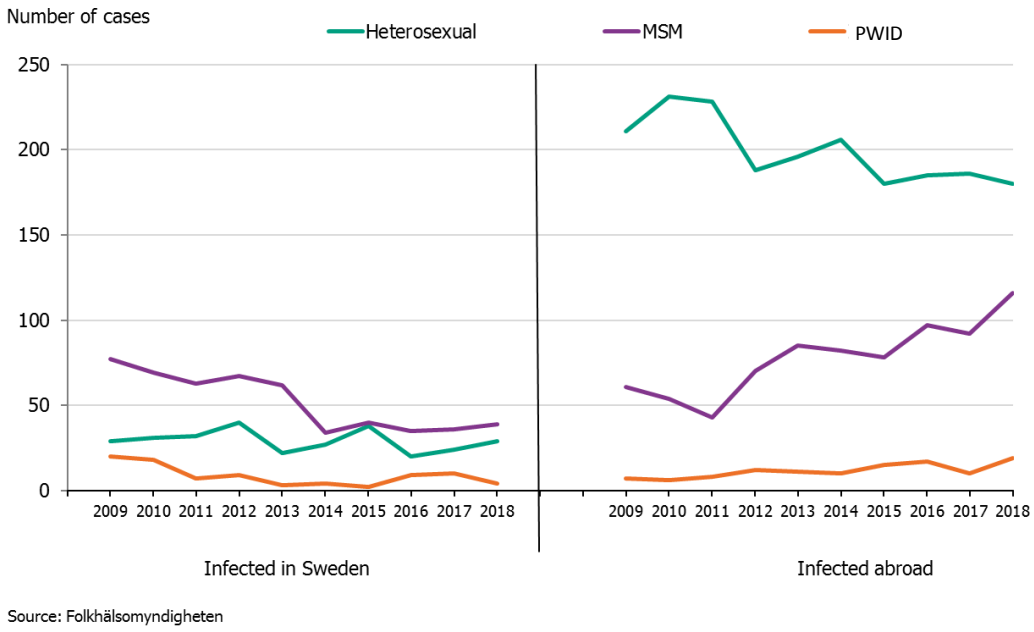

Fig 7. Reported HIV cases per route of transmission and country of infection, 2009-2018. Reprinted with permission. [70]

Preventive measures including ART have been successful in decreasing HIV-1 transmissions in Sweden. However, there is still transmission in the MSM group within Sweden and even more frequently during travel. Pre-exposure prophylaxis (PrEP) was recently introduced to strengthen prevention in MSM at high risk of HIV infection. Few endemic transmissions among PWID have been reported the last years, and harm reduction programmes including needle exchange are established in the major cities. [77]
Sweden is a resource-rich country with tax financed health care. All HIV care, including ART prescription, is free of cost under the Swedish Communicable Diseases Act. The legislation also supports extensive partner tracing around new cases of HIV-1, and demands individuals with HIV-1 to stay linked to HIV care. [78]
Sweden was reported to be the first country to reach the 90-90-90 target in 2017. [68] Analysis of the last two "90s" is reliable and based on the national database InfCareHIV. In 2018, 98% of diagnosed individuals were on ART, and 96% of those were virally supressed. (Personal communication, Veronica Svedhem) However, there is still an uncertainty in the first "90" that is based on estimates of the proportion of undiagnosed people living with HIV-1 in Sweden. The "fourth 90", on quality of life in PLHIV is endorsed by the steering committee of the national database InfCareHIV, and the first goal is that 90% of patients should complete a standardized form on quality of life. (Veronica Svedhem, personal communication)

A remaining issue in Sweden is late HIV-1 diagnosis and missed testing opportunities. In 2009-2012 58% of individuals were late presenters, with CD^+ T-cell counts <350/ul or AIDS at diagnosis, and this characteristic was associated with origin in SSA, Eastern Europe, Asia and the Pacific region. Of late presenting migrants, 66% had not been tested at immigration and half had lived in Sweden for more than a year at diagnosis. [79]

HIV-1 transmissions related to migration and travel in combination with heterosexual and MSM spread within the country has resulted in a high genetic variation among HIV-1 in Sweden. Starting from a subtype B-dominated epidemic, non-B subtypes have emerged over time and today Sweden has one of the most genetically diverse epidemics outside central Africa with presence of all HIV-1 M group subtypes, 17 CRFs and 32 URFs in an analysis from 2014. [80] With NFLG sequencing even more URFs were detected, and transmission clusters of URFs in MSM in Sweden were discovered. A significant increase of URFs in Sweden was shown from both *pol* and NFLG analysis. [27] A retrospective cluster analysis of HIV-1 transmission in the Nordic countries show that 85% of active networks and large clusters were subtype B clusters, indicating the continued importance of subtype B for local transmission. [34] This makes studies on HIV epidemiology in Sweden complex and highlights the importance of continued investigation of transmission networks in key populations and migrants.

1.2.4 Estimating undiagnosed populations living with HIV-1

Undiagnosed individuals with HIV-1 are at risk of progressive disease, AIDS and death and risk to transmit the virus to others. The UNAIDS 90-90-90 treatment target underscores the need to identify undiagnosed individuals to stop the spread of the virus and to treat everyone in need. Reliable estimates of undiagnosed populations are necessary to measure progress towards the target and to design and evaluate interventions. However, this is a challenge since general seroprevalence data is scarce in most settings. Mathematical modelling is now widely used to produce estimates of aspects of the treatment cascade both globally and nationally. [2, 67]

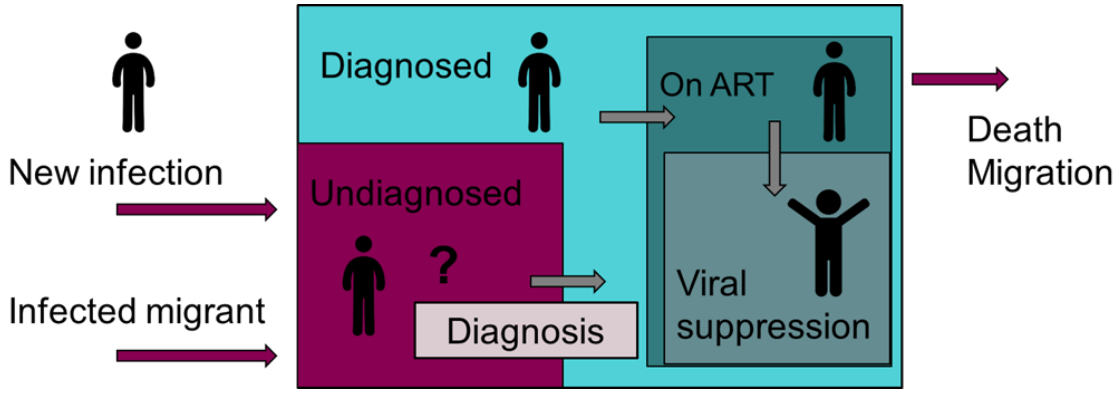

Fig 8. Visualisation of the concept of undiagnosed PLHIV (burgundy square) as a part of the total of PLHIV (large turquoise square) in a country.

The first attempts to use mathematical approaches to increase the knowledge of the HIV-epidemic were made in the late 1980´s when "back calculation" was introduced to estimate previous HIV incidence from AIDS incidence data and estimates of the distribution of the incubation period to development of AIDS. The method was used to assess the total number of infected individuals and the undiagnosed proportion. [81] With the introduction of ART that halted progress to AIDS, the method led to underestimation of the infected population. [82] Different statistical approaches were undertaken but there were still inherent problems in the method; it assumed a closed population, good quality AIDS-surveillance data and since AIDS incidence lags behind, estimates for recent periods were unreliable. [83]

As the methods shifted focus from AIDS incidence data to HIV case reporting, the need for determination of the true date of infection emerged. [84] "Extended back-calculation" was developed and integrated with the BED assay, a serological method to classify recent and long-term infections, with HIV case reporting and data on AIDS defining events at diagnosis. This approach was independent from effects of ART but suffered from the disadvantages of the BED-assay (misclassifications of some longstanding infections as recent and subtype-dependent results) and was dependent on reliable reporting. [85]

Based on the abundance of surveillance and cohort data in the United Kingdom, the HIV Synthesis model of HIV progression and the effect of ART was developed. It is an individual-based stochastic simulation model that simulates the course of infection in HIV-positive individuals from the start of the epidemic and forward, based on population level data including new HIV diagnoses, $CD4^+$ counts and ART-effects including possible drug resistance. Fictive individuals are modelled from infection (transmission is not modelled) and onwards, with parameters such as diagnosis status, $CD4^+$ counts, ART adherence and AIDS diagnosis being updated every three months in the model. Model outcomes are calibrated to observed data. General disadvantages are that it is mathematically advanced and computer intensive. [86, 87]

The back-calculation approaches were used mainly in high-income countries, where reliable data on AIDS diagnoses over time was available. For low-income settings, demography based models with less data requirements have been a way forward. [88]
The method currently used by most countries and recommended by WHO as default for estimates of the total number of PLHIV is the Spectrum AIDS Impact Model. The Spectrum program, first put together in 1995, is based on a demographic projection model that projects the population by age and sex and has multiple uses in examining trends and interventions related to reproductive health. Modules related to factors that influence demography are added on, including the AIDS Impact Model. The model and the assumptions, that reflect current knowledge of different aspects of HIV infection and ART, are updated regularly under the guidance of the UNAIDS Reference Group on Estimates, Modelling and Projections. [67, 89, 90] The main input to estimate HIV prevalence and the number of PLHIV in generalized epidemics is the prevalence in pregnant women at antenatal care centres combined with data from household/key population surveys and ART- and prevention of mother-to-child transmission (PMTCT) programmes. In epidemics concentrated in key populations, serosurveys in these populations and estimates of their size, or HIV case surveillance in combination with AIDS-related mortality is used. Spectrum estimates are considered robust, and data quality and coverage can be considered in the process. Uncertainties of the estimates are displayed as 95% plausibility bounds. The limitation is that as for any model, input data must be representative for the modelled population to produce reliable estimates. [2, 67, 91]

To address the need to make estimates of the care cascade including the undiagnosed proportion of HIV in Europe, the Stochastic Simulation of Outcomes of People with HIV In Europe in EuroCoord (EuroCoord-SSOPHIE project) was formed in 2011. The HIV-1 epidemics in European countries are diverse, but often concentrated to key populations, and with important influence of migration. The goal was to develop an improved individual-based model of HIV to make estimates of European PLHIV based on surveillance and clinical data. [92]. The resulting model (presented more in detail in the methods section 3.3.2) is a development of the HIV Synthesis progression model and is designed to model key populations and other risk groups separately and attempts to model the effect of migration from high-incidence countries in SSA. [87, 93]

European Centre for Disease Prevention and Control (ECDC) has designed a HIV modelling tool (described more in detail in the methods section 3.3.3) that is under continuous development and accessible without cost. An incidence model requiring less data than the above described and a back-calculation model only requiring the number of HIV/AIDS diagnoses stratified due to $CD4^+$ T-cell counts during at least one year, are provided. [94, 95]

1.2.5 Estimates of undiagnosed populations in Sweden

The first estimates of undiagnosed PLHIV in Sweden were based on back calculation and made by the Swedish Institute for Infectious Disease Control (now Public Health Sweden) in 1998. The estimated number of undiagnosed MSM in 1986 were 400-500 individuals, as many as the diagnosed patients. The number of undiagnosed individuals was estimated to have declined to no more than 200 in the 1990´s. [73] A study of the Nordic countries 1977-1995 using HIV and AIDS surveillance data in a back calculation based model estimated the proportion of undiagnosed HIV infections in MSM in Sweden to be 19% in 1995. [96]

When the Synthesis model (described in 1.2.4) was applied to European surveillance data from 2005 the proportion of undiagnosed PLHIV 2006 within EU was estimated to be 30%, 50% in the whole WHO European Region and the proportion in Sweden was estimated to 12-20%. [97] A new analysis based on surveillance data 2003-2015 and using the incidence method of the ECDC HIV modelling tool, estimated that 15% of PLHIV in Europe were not diagnosed in 2015, but no country-specific estimates were made. [98]

Our efforts (**Paper III**) to estimate the proportion of undiagnosed PLHIV in Sweden 2013 with established methods resulted in unexpectedly high estimates: 26% (n=2,100; 90% plausibility range [PR] 900-5,000) with the SSOPHIE method and 21% (n=2,013; 95% confidence interval [CI] 1,831-2,189) with the incidence method of the ECDC HIV modelling tool. Our conclusions were that we need better data on migration and further development of methods to obtain reliable estimates of the diverse Swedish HIV-1 epidemic. [99]

The Swedish Public Health Agency reported an estimate of 10% undiagnosed patients 2015 based on surveillance data, that supported the achievement of the 90-90-90 target in 2015. [68] This preliminary estimate was produced by back-calculation from surveillance data and serosurveys in blood donors. However, blood donors are not representative of the general population as they are selected to have low risk for blood-borne infections. (Maria Axelsson, personal communication)

A novel approach relies on biomarker-based estimates of time to diagnosis in newly diagnosed individuals and can discriminate between infections within and outside the country by adding migration and surveillance data to a Bayesian model. This method estimated the undiagnosed proportion of PLHIV in Sweden to be 816 individuals in 2015, a number that equals 10.8% (95% CI 10.3-11.4%) of PLHIV. [100]

1.2.6 Estimating time to HIV-1 diagnosis

The long clinical "latency" period from infection to symptomatic disease in HIV-1 infection often makes it hard to assess when a patient was infected. In most cases, anamnestic and clinical information is not enough. Laboratory confirmed primary HIV-infection and seroconversion data with narrow intervals are the most reliable data but are sparse. Patients with low $CD4^+$ T-cell counts are more likely to have been infected for a long time, but since the rate of $CD4^+$ decline is variable this is not completely reliable. [101] During primary HIV infection the $CD4^+$ T-cell counts are most often low during the first month and pose a risk of misclassification of the infection as long-standing. [102]

To estimate HIV-1 incidence and for higher accuracy in addressing undiagnosed PLHIV, good estimates of time to diagnosis are helpful, and although much effort has been put into this field of research, no method is yet reliable at the individual level.

Algorithms based on serological assays have been widely used for epidemiological purposes and can discriminate between recent and chronic HIV-1 infection. [103, 104] There are three different approaches of serological assays; measurement of the level of HIV-1 antibodies ("detuned assays", not widely in use today), measurement of the proportion of HIV-1 specific IgG related to the total IgG (BED capture enzyme immunoassay), and measurement of avidity of HIV-1 antibodies (Limited Antigen Assay (LAg)). A disadvantage of the BED assay is that it misclassifies a proportion of long-term infections as recent and that a calculation of false-recent rates specific to the studied population is required. Furthermore, HIV-1 subtype influences performance and this is problematic in non-subtype B and epidemics with mixed subtypes. Usage of BED results can be maximized with addition of a mathematical model that allows assessment on a longitudinal scale instead of just a dichotomous result. [105] LAg assays are more sensitive but can be affected by tuberculosis infection and misclassify HIV-1 subtype D. [106-108] Advantages of serological methods are that they are relatively cheap and can be made on serological samples or dried blood spots without any additional data required. [109] High false recent rates in individuals on ART can be reduced by introducing viral load as an additional parameter. [110]

An alternative approach is to use the increase of viral diversity during infection as a marker of time since transmission. In *pol* sequences collected for routine genotypic resistance testing by Sanger sequencing this is represented by an increase in the proportion of ambiguous (mixed) base-calls. The method can discriminate between recent and long-term infection in various subtypes if a dichotomous cutoff is used, but weaknesses are misclassification of long-standing infections with decline in diversity, and that it does not work well in individuals infected with more than one viral variant. Methodological variations such as different thresholds for calling mixed bases need to be considered. [111-113]

Next-generation sequencing/High-throughput sequencing (NGS/HTS) with deep sequence analysis of viral variants can assess diversity with higher accuracy, and potentially overcome the obstacle of multiple founder strains in recent infection. [114-116] This approach gives

possibility to assess the time of infection on a continuous scale rather than dichotomous, increasing the usefulness beyond incidence assays. The increased sensitivity is based on detection of minority viral variants below the detection limit of Sanger sequencing. Different parts of the genome have been used for assessment of viral diversity. [114, 116] *Puller et al.* found an optimum in 2,000-3,000 bp of *pol*, with a linear increase of viral diversity in the third position of the codon, at least 8 years after infection. The precision of estimates was +/- one year in long-time infection and slightly lower in more recent infection. [116] This was corroborated on a larger material (n=313) from the Swiss cohort, although mainly subtype B sequences. A direct comparison to mixed base-calls in Sanger sequences was made and clearly stated the advantage of the NGS-based analysis. [114] Another study chose the *env* gene and integrated the approach with an algorithm to calculate HIV incidence. [117]

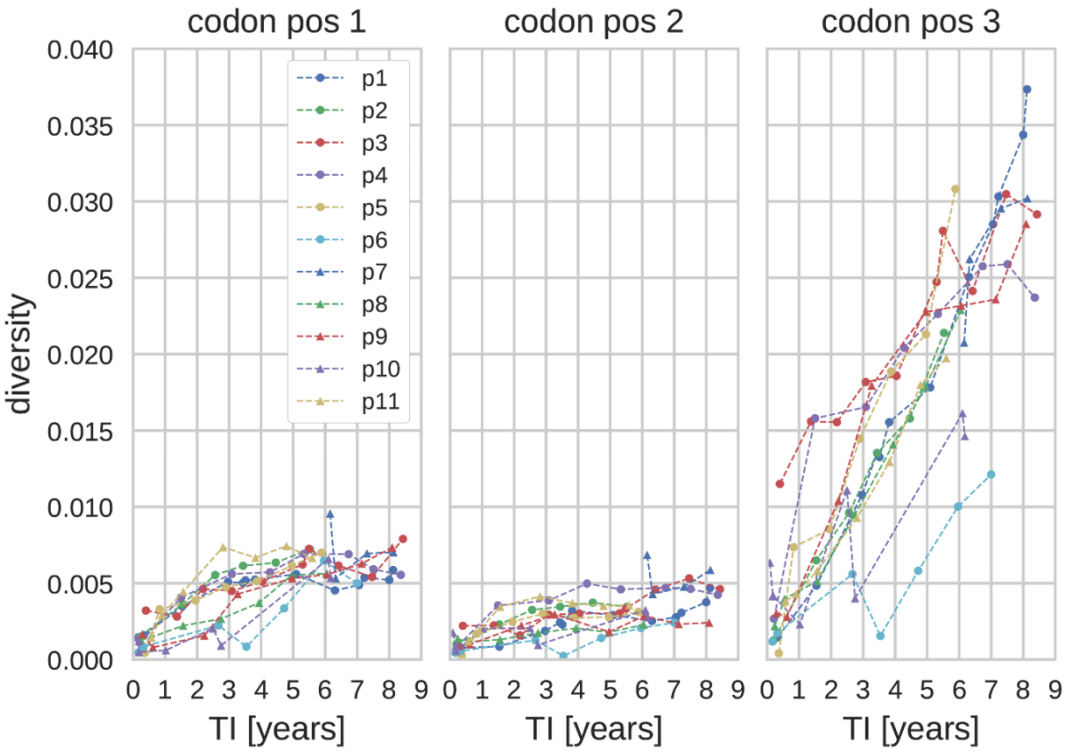

Fig 9. Diversity in *pol* (average pairwise distance) as a function of time since infection (TI) in eleven individuals with well-defined time of infection and sequential naïve samples. Reprinted with permission. [116]

Even though $CD4^+$ T-cell counts at diagnosis do not suffice for estimating time from infection, they carry useful information and performance of estimates can be improved by mathematical modelling. A $CD4^+$ T-cell decline trajectory model that is based on modelling of the $CD4^+$ T-cell slope from infection to diagnosis with consideration of effects of patient age and region of origin on $CD4^+$ T-cell decline has been developed. [74, 118]

To improve estimates and address weaknesses of individual assays and biomarkers, algorithms combining parameters have been suggested. [100, 108, 119] A recent publication showed that the optimal combination for incidence estimates in subtype C in African women

was a combination of two serological avidity assays and VL. However, viral diversity was not tested as part of the algorithm. [120]

So far there are no published data on combination assays including viral diversity measured with NGS. Usefulness of algorithms depends not only on their performance, but also on availability of data. With increasing use of NGS for GRT worldwide, estimates based on viral diversity will be more accessible. Even though NGS estimates viral diversity with high resolution, the issue with multiple founder viruses remains [114], but can potentially be addressed with serological assays, that have discriminatory strength in early infection.

1.2.7 Estimating country of HIV-1 infection

In settings where migration and travel are important factors for HIV-1 epidemiology, it is important to distinguish between domestic and foreign transmissions. Both local infection control interventions and modelling efforts of HIV seropositive populations depend on correct information on country of infection. In 2017, 41% of individuals reported with HIV infection in the EU/EEA were not born in the reporting country. [121] Sweden reports the highest proportion of migrants (75% 2018) among newly diagnosed individuals in Europe. [70] A systematic review of post-migration HIV acquisition in Europe underscores the clinical and epidemiological relevance of HIV infections occurring in migrant groups in their new country of residence, and the need for surveillance and prevention initiatives designed for these populations. [76] Infections acquired post-migration are not necessarily acquired in the country of residence, but still possible to prevent by targeted national initiatives.

To collect reliable data on country of infection, doctors´ reports are not sufficient, since they tend to underestimate infections after arrival. To improve estimates, combinations of methods to estimate time to diagnosis, collection of dates of arrival and phylogenetic approaches have been attempted. [74, 75, 118]

The French PARCOURS study assessed that the proportion of post-migration acquisition of HIV in PLHIV from sub-Saharan Africa living in France was at least 35%, based on real-life data in combination with $CD4^+$ T-cell modelling. [122] A British study using $CD4^+$ T-cell modelling tripled the estimate of domestic infections in foreign born to 33% compared to the estimates by clinical reports (11%). [118]

A study on migrants in Sweden diagnosed with HIV-1 during 1983-2013 estimating time to diagnosis with a $CD4^+$ T-cell decline trajectory model and comparing the results to date of arrival, also suggested that domestic infection in migrants was more common (19%) than indicated by doctor´s estimates (12%). Further analysis with phylogeny of *pol* sequences corroborated the results. [74]

The aMASE study used a Bayesian approach on epidemiological data integrated with $CD4^+$ T-cell counts and viral load trajectory data on migrants diagnosed with HIV in nine European countries. In this cohort, acquisition of HIV post-migration was 63%, with the highest proportion, 72%, in MSM. Individuals with origin in Latin America or the Caribbean were

infected after arrival in 71%, and 45% of migrants from SSA were infected post migration. [75]

1.3 HIV-1 ANTIRETROVIRAL THERAPY

1.3.1 Development of antiretroviral therapy

One of the first antiretroviral therapy trials ever performed was with the drug foscarnet at Roslagstulls Hospital, Stockholm, in 1985-1986. Later it was shown that foscarnet had a pronounced antiviral effect on HIV-1 with declining levels of HIV antigen [123] and HIV RNA during therapy. [124] Also novel approaches like a small peptide [125] and alpha-interferon were tested. [126] Antiretroviral therapy for HIV-1 on a larger scale was first introduced in 1986 with the nucleoside analogue zidovudine (AZT), and reduced mortality significantly during short-course studies [127-129]. However, the effect on survival decreased after 15-18 months of therapy and in 1989 a phenotypic assay revealed AZT-resistance in virus isolated from patients treated for more than six months [130]. The same pattern of mutations in the RT was found in all resistant strains, and evidence of the sequential accumulation of the mutations was presented. [131]

The need for combination therapy for HIV-1 to avoid development of resistance was recognized early, based on the properties of the virus and empirical evidence. Through studies of the HIV-1 life-cycle potential targets for antivirals were identified and drugs targeting different steps in the viral replication were developed during the early 1990´s; additional nucleoside RT inhibitors (NRTI), non-nucleoside RT inhibitors (NNRTI) and protease inhibitors (PI). However, no drug or drug combination succeeded in durably suppressing VL, and signature mutations for respective drug and/or drug class were detected in resistant strains. [132, 133] It was shown that sequential therapy favoured development of multidrug-resistant virus both *in vitro* and *in vivo*. [134, 135]

In 1996, a combination of the most potent drug class, PI, with two NRTIs, proved to be effective in durably reducing VL and mortality in AZT-experienced patients. [136-138] Another breakthrough was the identification of HIV-1 RNA quantification in plasma as a valid marker for treatment efficacy and prognosis for patients on treatment. [139, 140] Accumulated resistance mutations leading to multi-drug resistance were a major issue in patients with treatment experience. The need for pre-treatment and treatment-failure GRT was identified, and with technical achievements sequencing of viral genes was suggested as a feasible method. [141]

In 1996 a standard-of-care with two NRTIs combined with a PI or NNRTI backbone was established. Viral suppression under 500 copies HIV-1 RNA/ml was identified as a treatment goal but still not reachable in all cases. [142-144] The addition of ritonavir-boosting of PIs [145] and further development of antivirals has resulted in modern ART regimens that suppress VL to undetectable levels with a low pill-burden and much less side-effects. With the development of more sensitive HIV RNA quantification assays the cutoff of

"undetectable virus" has now been moved to 20 or 50 copies/ml, depending on the assay. [146] Today, seven classes of antivirals targeting different parts of the HIV-1 life cycle are available; NRTI, NNRTI, PI, integrase strand transfer inhibitors (INSTI), CCR5-inhibitors, fusion inhibitors and post-attachment inhibitors. The three last drug classes inhibit entry into the cell and are seldom used. A new class, the capsid inhibitors, is in early clinical trials and has shown high potency and promises to be useful as a long-acting drug. [147]

INSTI are an increasingly used alternative to NNRTI and PI since registration of the first compound of this class in 2007. Besides a rapid decline in VL at treatment initiation, INSTI also have the benefits of a low frequency of side effects and drug-drug interactions. [148] However second generation INSTI dolutegravir (DTG) and bictegravir (BIC) are still under investigation for association with an increased frequency of neural tube defects in pregnancy [149] and weight gain has been associated with DTG. [150]

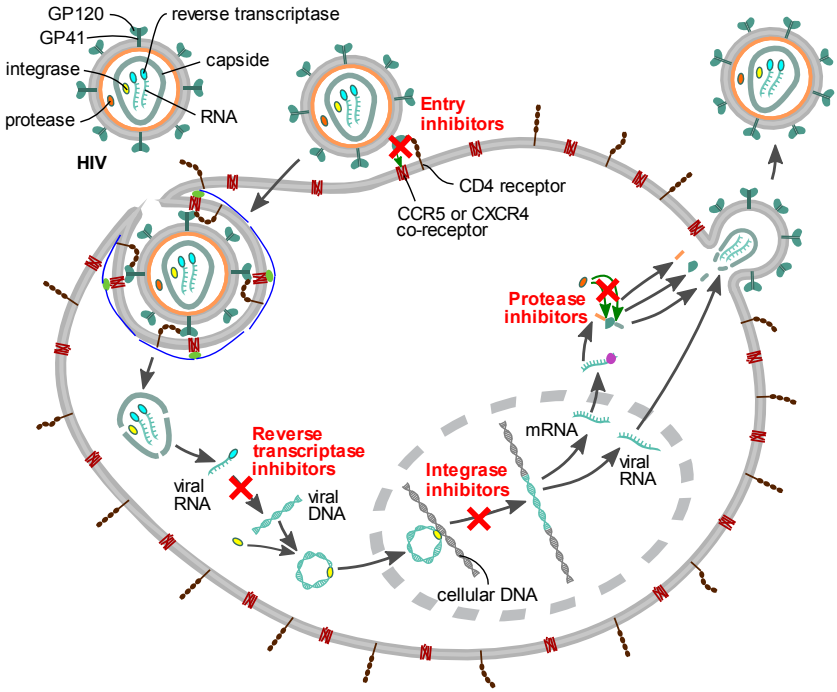

Fig 10. Schematic presentation of the impact of different classes of antivirals on the HIV-1 replication cycle. *Wikimedia Commons. Thomas Splettstoesser* (www.scistyle.com)

1.3.2 Monitoring of treatment

Quantitative HIV-1 RNA measurements in plasma as a measure of viral suppression and $CD4^+$ T-cell counts in blood to monitor immune reconstitution have proven individual and combined value in predicting outcomes in PLHIV on ART. [140, 151] Effective ART usually reduces the VL in plasma below detectable levels of sensitive assays (<20 or <50 copies/ml) in 8 to 24 weeks, and maintains viral suppression thereafter.

Insufficient drug effects due to antiviral resistance or suboptimal drug concentrations can lead to virological failure, commonly defined as repeated HIV-1 RNA measurements >200

copies/ml. Reasons for insufficient effect of modern ART regimens are suboptimal adherence, decreased uptake from the gut or drug-to-drug interactions. Identification of HIV RNA in plasma during ART should lead to evaluation of the current regimen, including adherence, and the possible need for regimen switch and/or GRT. [146, 152] Continuation of a failing regimen will lead to accumulation of resistance mutations and further loss of regimen efficacy, also in drugs with a high genetic barrier such as DTG and boosted PIs. [153, 154] A meta-analysis has shown that individuals with virological failure that had not been frequently monitored with VL had more resistance mutations at time of failure than those who were closely monitored. The probable reason is that delayed detection of virological failure allows for further development of resistance. [155] Continuous monitoring of VL during ART ascertains that individuals get the best treatment effect and that the risk of transmitting HIV-1 is negligible when undetectable levels of HIV RNA in plasma are achieved. [152] "Blips" are occasional events of detectable low-level viremia with subsequent return to undetectable levels in an individual with viral suppression, and are usually not related to virological failure or development of drug resistance, but rather reflect stochastic variations in actual VLs and VL measurement precision. [156] However, they can be predictive of later virological treatment failure. [157] There is still controversy on the significance of persistent low-level viremia 20-200 copies/ml, but evidence exists that resistance can emerge even at this levels under the selective pressure of ART. [152, 158-160] Viral blips and low-level viremia have been associated with slower decay of HIV-1 reservoirs under ART, something that may be of importance in HIV-1 cure efforts. [161]

1.3.3 Clinical impact of HIV-1 subtype

Several studies have suggested that HIV-1 subtype influences virulence and disease progression in naïve individuals, but the evidence is conflicting. Association of subtype with ethnicity, route of infection and other epidemiological factors is a potential bias. [162-166]

In VL monitoring and GRT, subtype-related viral diversity is a challenge due to variation in targets for PCR and sequencing. Many assays were originally designed for subtype B, resulting in non-quantification or sub-quantification of strains with differences in target sequences. Local subtype representation needs to be considered when introducing new laboratory assays, and in interpretation of conflicting laboratory results. [167-170]

Today, there are no subtype-specific recommendations on ART regimens. Potential subtype-related differences in efficacy of antiretrovirals are of major interest when extrapolating ART efficacy results from studies mainly on subtype B to settings where other subtypes are dominant.

A review of available studies showed small differences in individuals on ART but there were subtype-specific differences in resistance mutation patterns in treated populations. [171] In a Swedish study from 1998 - 2002, where most patients were treated with ART regimens with unboosted PIs, the outcome in individuals with African origin was inferior, but no relation to HIV-1 subtype was proven. [172] A recent comparative study in the Swedish InfCareHIV

cohort showed higher risk of both primary and secondary virological failure in subtype C than subtype B when boosted PI were used. Molecular dynamics structure simulation and docking analysis of the subtype C protease, suggested that the binding of PIs is weaker in subtype C due to naturally occurring polymorphisms and that this explains a weaker treatment effect. [173] A clinical study from US points in the same direction. [174] In addition, a PYxE insertion in Gag-p6 has been found to increase the viral fitness of subtype C strains and also be associated with an increased risk of failure on boosted PI therapy. [175-178] Results from a clinical study have also indicated that CRF 02_AG strains are inherently less sensitive to boosted PIs than subtype B strains. [179] Also second generation NNRTIs etravirine and rilpivirine have shown lower potency in subtype C strains in *in vitro* experiments. [180, 181]

INSTI have been introduced in low- and middle-income countries (LMIC) recently and real-world efficacy data on other subtypes than B are sparse. Indications of significance of subtype-related polymorphisms in combination with pre-treatment findings of major INSTI mutations in minority viral variants in subtype C strains from Ethiopia points to a need of further investigations of subtype-related differences in efficacy. [182-184] Failures on the long acting INSTI cabotegravir in a clinical study have recently been associated to sub-subtype A6 and possibly to the polymorphism L74M/I. [185]

1.4 HIV-1 DRUG RESISTANCE

1.4.1 Genetic barriers to resistance

All modern ART combination regimens block viral replication and maintain long-term viral suppression if drug concentrations suffice and no pre-treatment drug resistance (PDR) is present. However, regimens differ in robustness if adherence fails. [154] The risk of resistance development is decreased with drugs with high potency and a high genetic barrier. The antiretroviral drugs and drug classes have different barriers to resistance, depending on how many mutations in the viral genome that are needed to lead to clinically significant resistance and the consequences for the virus of the mutations. [160] So far, no drug class is robust enough to use for monotherapy, but regimens with two drugs have been suggested and are already in clinical use in some settings. However, the standard of care is still three antiretrovirals in combination. [186]

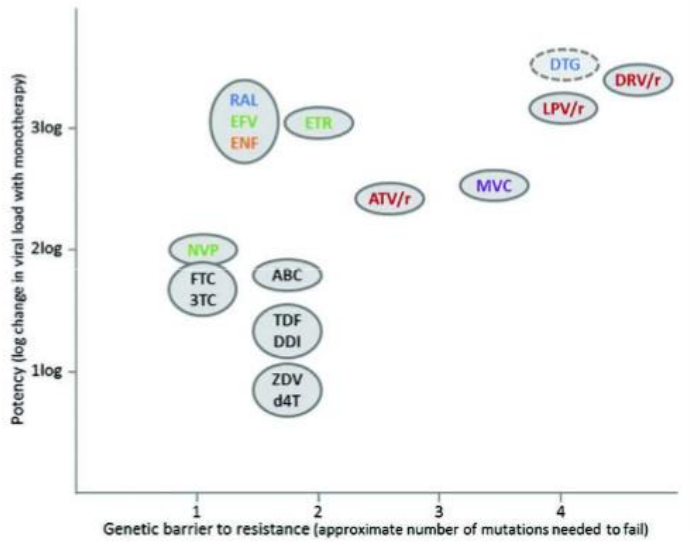

Fig 11. Potency vs genetic barrier to resistance in important antiretrovirals for HIV-1 treatment. NRTI (FTC, 3TC, ABC, TDF, DDI, ZDV(=AZT), d4T), NNRTI (EFV, NVP, ETR), PI (ATV/r, LPV/r, DRV/r) INSTI (RAL, DTG), Fusion inhibitor ENF and CCR5-inhibitor MVC. [160] Reprinted with permission.

In drugs with a high genetic barrier, resistance development requires several steps, where the virus has opportunity to replicate under insufficient, but still selective, concentrations of the drug for a long time. This favours selection of variants with drug resistance mutations, and in the following steps drug resistant variants may develop both increased resistance and compensatory mutations that increase viral fitness. Due to viral diversity, many single drug-resistance mutations are probable to exist pre-treatment by chance, but this is not the case for combinations of resistance mutations. [187]

In the example of boosted PIs, that generally have a high barrier to resistance, the resistance-causing "major" mutations usually decrease the fitness of the virus and are followed by "minor" mutations that restore the fitness of the resistant virus to make it more competitive. Minor mutations are frequently wildtype variants, i.e. polymorphisms, in some subtypes of HIV-1. [188-190]

Important examples of drugs with a low genetic barrier are the widely used first generation NNRTIs efavirenz (EFV) and nevirapine (NVP) and NRTIs emtricitabine (FTC) and lamivudine (3TC) as well as the INSTIs raltegravir (RAL) and elvitegravir (EVG). For these drugs, one mutation is enough to create a virus with high resistance. This makes these drugs particularly unforgiving to insufficient drug concentrations, since highly resistant variants will readily be selected once the virus is allowed to replicate under drug pressure. [191] The accompanying drugs are also endangered if the regimen is not switched promptly. [192]
For other NRTI drugs, the resistance is complex in that different evolutionary pathways towards high-grade NRTI resistance exist, and that patterns of mutations affect the individual NRTIs differently. [187] The two alternative pathways of thymidine analogue

mutations (TAM), induced by the early NRTIs AZT and stavudine (d4t) exemplify this. [193] Viral subtype is another factor that influences how resistance patterns emerge during virological failure, depending on differences in the original viral sequence. [194, 195] Recently a new RT inhibitor, 4-ethynyl-2-fluoro-2-deoxyadenosine (EfdA) has been found to have a higher *in vitro* genetic barrier to drug resistance that earlier NRTIs. [180]

INSTI are very potent drugs that give rapid viral decline. The first generation INSTI (RAL and EVG) has the disadvantage of a low genetic barrier [196], but the second generation, dolutegravir (DTG), bictegravir (BIC) and cabotegravir, were developed to have a high genetic barrier. [197] As in the case of PIs, several evolutionary steps (mutations and selection) are required for development of resistance. However, monotherapy has failed in clinical trials and at least one potent companion drug is necessary. [198]

1.4.2 Acquired drug resistance

Acquired drug resistance in HIV-1 infection may develop when viral suppression is insufficient in an individual on ART, usually due too poor adherence or pharmacological interactions. Minority viral variants harboring resistance mutations will then appear due to *de novo* mutations and can be selected by evolutionary pressure. Their ability to replicate in the presence of ART will lead to their enrichment in the major viral population. Drugs with lower genetic barriers to resistance are usually affected first, and by switching to second line treatment promptly, resistance to accompanying drugs may be prevented. Prolonged viral failure may lead to multiclass drug resistance. [199, 200]

In many cases, cross-resistance is of clinical importance, meaning that other drugs of the same class are compromised by the acquired mutations. Cross-resistance is not always absolute and differs between drugs and drug classes. In drug development (i.e. NNRTI and INSTI) care has been taken to try to avoid cross-resistance between first generation drugs and following compounds. Within the NRTI-class, a mutation can decrease or increase antiviral effects of different NRTIs (i.e. M184V reduces sensibility to 3TC and FTC but increases the effect of tenofovir (TDF)). [160, 201]

The viral subtype has some impact on the development of certain drug resistance mutations, for example V106A is more common in subtype B and V106M in subtype C, in response to failing NNRTI regimens. Subtype C has a higher tendency to develop K65R that confers NRTI-resistance, especially to TDF. [160]

The integration of resistant proviruses in the genome of $CD4^+$ T-memory cells creates an "archive" of resistance mutations and "treatment/resistance history" needs to be considered in individuals restarting or switching ART regimens. [200, 202]

1.4.3 Transmitted drug resistance

Drug resistant variants of HIV-1 may be transmitted from a patient with acquired resistance, and further on from one naïve person to another. Transmitted drug resistance

(TDR) in HIV-1 was first described in 1993 [203-206], but anticipated already when the first AZT-resistant viral strains were found [131]. TDR can hamper the success of first-line ART if not accounted for in regimen design. [207, 208]

The continuous reshaping of the viral quasispecies leads to reversion of less fit resistant viral variants after transmission to a naïve individual. If they only exist in minority viral variants they will not be detected by routine population based sequencing but may still be of clinical importance. Depending on their impact on fitness, there is variation in persistence of different drug resistance mutations in population-based sequences. A cohort study of ART naïve individuals with TDR showed that among NRTI-mutations, M184V reverted quickly with a median time of one year, while M41L and other TAMs were detected for much longer. Mutations in position 215 did revert, but the revertant variants were conserved. Reversion of NNRTI mutations was faster than for the TAMs, K103N had a median time to loss of 3.7 years. [209]

A consensus cut off ≤180 days from diagnosis has been used for sampling for resistance testing in TDR studies to increase sensitivity, but the date of infection, when reversion probably starts, is seldom known.

In vitro studies show that some resistance mutations decrease the replicative fitness of the virus (e.g M184V, K65R and T215Y), but others do not affect fitness as much (e.g L90M, Y181C, D67N, L210W). Whether the same effect on transmission fitness applies has been uncertain. [210] Several approaches have been applied to specifically study transmission fitness; comparisons of mutation prevalence in treated and untreated patients from the same epidemic [211], relative tendency of strains with and without resistance mutations to form clusters in a molecular transmission network [210] and a phylodynamic approach that assessed transmission rates of wild type versus resistant strains. [212] There is consensus from at least two of these studies on that the NRTI mutations K70R, M184V and T215Y have lower transmission fitness, that M41L, the T215 reverts, and the important NNRTI mutation K103N do not hamper transmission and it is also indicated that the PI mutation L90M might increase transmission fitness.

In Europe and the US, transmitted TAMs and NNRTI-mutations are frequently related to clonal spread from ART-naïve individuals. [210, 213-217]

In contrast, in SSA TDR has been suggested to be more closely related to spread from individuals on failing regimens, since clustering patterns are uncommon in studies. [218] However, sampling of a small fraction of individuals could underestimate spread of TDR from naïve individuals. [217] Findings from a recent study in Ethiopia support that TDR from naïve individuals is important also in SSA; a phylogenetic sub-cluster with the NNRTI mutation G190A was traced from 2003 – 2010. [219]

So far, transmitted resistance to INSTI is uncommon, and baseline integrase GRT is not generally recommended in most guidelines although this is routinely performed in certain

areas such as France and Stockholm (Personal communication, Anders Sönnerborg). However, case reports indicate that major INSTI mutations E138A, G140S, Q148H and N155H can be transmitted and the first report of transmitted high-level resistance to all INSTIs was recently published. [220-222]

1.4.4 Pre-treatment drug resistance

Pre-treatment drug resistance (PDR) is any drug resistance in individuals about to initiate first-line ART. In most settings this means a combination of individuals who are newly diagnosed with HIV-1 and of those who have been on treatment before but stopped. PDR measures thus a combination of TDR and acquired drug resistance. It is still important to differentiate the types of drug resistance, but lack of medical history often makes it hard to differentiate TDR from acquired resistance, especially in migrants. PDR is a pragmatic concept that mirrors the clinical resistance situation. [199, 223]

1.4.5 Resistance testing

Testing for HIV-1 drug resistance (HIVDR) can be performed by phenotypic or genotypic methods.

Phenotypic testing assays are technically more demanding and more expensive. They measure the growth of HIV-1 in a cell culture under different concentrations of a certain antiviral, to determine the median inhibitory concentration (IC_{50}) and the fold increase of IC_{50} between the tested strain and a sensitive reference strain. When phenotypic assays were first developed, HIV-1 isolates from a patient were cultured, leading to a risk of the results not being representative of the *in vivo* strain. Replacement with recombinant assays where a *pol* sequence from the plasma virus is inserted into a laboratory HIV-1 strain, has improved the technique. These methods are best applied in studies of effects of new mutations/new drugs or in extremely complex clinical situations as a complement to GRT. Phenotypic testing of new generations of drugs renders vital information on cross-resistance due to mutations selected by compounds from the same drug class. [199, 224]

GRT to NRTIs, NNRTIs and PIs at baseline and at virological failure is an integrated part of HIV-1 care in high income settings. The routine method is still mostly population based (Sanger) sequencing of *pol* with commercial assays or *in-house* methods that target a *pol* fragment, including relevant parts of RT (for NRTIs and NNRTIs) and protease (for PIs). The sequence is then analyzed for presence of known drug resistance mutations (DRM). Resistance testing for integrase requires sequencing of an additional fragment of *pol* and international guidelines only recommend it in failure on INSTI regimens and if TDR to other drug classes is found. [199]

A general limitation of Sanger sequencing is that only resistant variants representing approximately 20% of the viral population are detected. [37, 225-228] With the advent of NGS/HTS viral variants down to 1% can be reliably determined, increasing sensitivity of detection of drug resistant variants that constitute a minor proportion of the virus population

but potentially decreasing the clinical specificity of findings. So far there is some evidence that NNRTI mutations down to 5% may be clinically relevant, since presence of minority viral variants of NNRTI-resistant virus has been reported to increase the risk of virological failure. [229-231] For the other drug classes a conservative reporting of variants 15-20% is suggested since there is no evidence of a benefit of reporting lower proportions of resistant variants. [229, 232] This is suggested by, among others, *Inzaule et al.* who base their recommendations on a large, prospective case control study in SSA. [229] In contrast, a study by *Derache et al.* from South Africa showed that NNRTI mutations alone had no effect on time to viral suppression on efavirenz-based ART neither above the 5% nor 20% threshold. However, NRTI and NNRTI mutations in combination prolonged the time to viral suppression when a 5% detection level was applied. The discrepancy to other studies is suggested to depend on the higher potency of a TDF/FTC containing backbone than of the thymidine analogue backbone that dominated in other cohorts. [233] A Swiss study demonstrated that viral suppression was obtained in individuals on boosted PI and 2 NRTIs despite presence of the NRTI-mutation M184V in minority variants. [234]

A practical advantage of NGS/HTS resistance testing is that a larger fragment of 3,000 bp can be sequenced and therefore only one test is required to also obtain INSTI resistance results. So far there is no convincing evidence that pre-treatment minority viral variants with INSTI drug resistance mutations are of clinical importance although very little data are available. [235]

NGS/HTS resistance assays generate massive amounts of data and a bioinformatic pipeline is needed to process sequencing data and produce resistance reports for clinical purposes. Several non-profit and commercial pipelines handling HIVDR HTS data are available, with varying demands of bioinformatic expertise for the user. In 2018 a multidisciplinary meeting resulted in *"the Winnipeg consensus"* with recommendations to harmonise HIVDR testing with NGS. [232]

Most evidence for associations between DRMs and clinical outcomes is generated from plasma samples, hence that is the material of choice. [199] However, dried blood spots (DBS) have been successfully used in LMICs and facilitate handling and shipping of samples to centralized laboratories since no cold chain is required. [236] Also a "split" procedure has been suggested where the amplification from plasma is done locally and the product sent to a central laboratory for sequencing. [237] To study archived resistance mutations in individuals with complex treatment histories, sequencing of proviral DNA in peripheral blood mononuclear cells (PBMC) has been suggested, but the clinical usefulness remains to be proven. [160]

The analysis of sequences generated for resistance testing is complex and the most common method is to apply a rule-based algorithm such as the Stanford HIV Drug Resistance Database, French National Agency for Research on AIDS and Viral Hepatitis, Rega, or HIV Genotypic Resistance-Algorithm Deutschland. In bioinformatic pipelines for NGS/HTS HIVDR, an algorithm needs to be part of the pipeline. The algorithms integrate information

from *in vitro* experiments, sequencing of HIV-1 from individuals on failing regimens and clinical studies with expert opinions to formulate a set of rules on the effects of certain mutations and combinations of mutations. The reports produced state susceptibility or level of reduced susceptibility to the commonly used drugs. [199, 238] The IAS-USA Drug Resistance Mutations Group compiles new data in a list that is updated regularly. [189]

Machine learning approaches are an alternative to the rule-based algorithms. [199, 239-244] They have the potential of integrating the genotypic data with other parameters of relevance for treatment success and hence produce more individualized predictions but are not widely used in the clinic today. In a head-to-head comparison between a predictive engine and "resistance experts", the machine learning was found to be a better predictor of which drugs to choose after a virological treatment failure than the experts. [245]

To implement resistance testing for clinical purposes in LMICs, where sequencing facilities are limited, point-of-care testing with rapid reporting of the most important mutations has been suggested. A major challenge in design of such assays is the sequence variation surrounding the drug-resistance positions of interest. [246, 247] Development of methodology for point-of-care testing of specific HIVDR mutations include PCR based approaches such as oligonucleotide ligation assays, multiplex allele specific assays and pan-degenerate amplification and adaptation, but also methods based on isothermal amplification. Implementation of any assay as part of ART programmes in resource limited settings is dependent on adequate training in handling and interpretation of the test, quality assurance programs and a functioning infrastructure including reliable electricity. [248]

1.4.6 The global perspective

In high-income settings, ART regimens are individually designed to have a low pill-burden, minimized side-effects and the choice of regimen is guided by baseline GRT to detect PDR in PR/RT, increasingly often also in IN. Therapy is monitored with HIV-1 RNA quantification and $CD4^+$ T-cell counts, to timely detect virological failure and address adherence issues or emerging drug resistance. PDR can be handled clinically although pre-exposure (PrEP) and post-exposure (PEP) prophylaxis to HIV-1 may still be compromised.

In low- and middle-income settings on the other hand, an individualized approach is usually too costly and logistically challenging to reach all in need of treatment. [154] The public health approach to ART was introduced by WHO in 2006 and made roll-out of treatment feasible by simplified and harmonised ART regimens. A standardised first-line regimen with EFV (NNRTI), TDF and FTC (NRTIs) has been recommended to all adult patients from 2016. VL monitoring was introduced in WHO guidelines in 2016 but is still not widely available or has long turn-around time from sampling to result reporting. The disadvantage of an NNRTI-regimen, which has a low genetic barrier to resistance in combination with limited access to VL monitoring is that patients will stay on a failing regimen longer because switch to second line will be based on clinical or immunological parameters, and viral variants with more complex resistance will be selected. [237] Also, an increasing proportion of individuals

initiating first-line ART in LMIC are not naïve, but rather re-initiating a NNRTI-based first-line treatment, and the treatment history may not be known to the health care facility. [249]

Up to 24% of patients on first-line treatment in SSA experience viral failure within 12 months. Resistance patterns are homogenous with standard first-line treatment; NNRTI-mutations and certain NRTI-mutations such as M184V (to FTC) and K65R (to TDF) are found. [250, 251] In the absence of pre-treatment resistance testing TDR results in worse treatment outcomes. [252] Availability of PI-based second line regimens is often limited in LMIC due to the cost being around three times that of the first line regimen, and less than 5% of individuals on ART in LMIC are on second-line regimens. [253]

The current WHO recommendations to defer switch to second line ART if VL is detectable but less than 1000 copies/ml, and to repeat VL testing in 2-3 months if VL is ≥1000 copies/ml before initiation of regimen switch have been questioned, as they may lead to development of advanced resistance and increased rates of TDR. A comparative study of switch strategies in a high-endemic, low resource setting in SSA is underway. [254, 255]

As a response to rising NNRTI resistance in LMIC, WHO now recommends the INSTI DTG as the preferred core agent in first- and second line ART. [256] Although DTG has a high genetic barrier to resistance, treatment success relies on effective companion drugs to avoid monotherapy that would inevitably lead to development of resistance. In this context, timely switch strategies are of importance.

1.4.7 Resistance surveillance and epidemiology

The local prevalence and nature of TDR is of importance to guide recommendations on first-line regimens, PrEP and PEP and is ideally studied by surveillance of resistance mutations in treatment-naïve individuals close to time of infection. To allow for comparisons between geographic areas and over time, a consensus list of Surveillance Drug Resistance Mutations (SDRMs) has been developed. The mutations on the list are unambiguously related to drug resistance, non-polymorphic, confer drug resistance in all common subtypes, and are not too rare in treated patients. [257] The list is integrated in the Stanford Calibrated Population Resistance online tool, facilitating analysis of large numbers of sequences. [258] Findings of SDRMs do not always confer clinically relevant resistance and require further analysis with one of the clinical algorithms (described in 1.4.5) as well as interpretation in the context of current ART regimens. [259]

In high-income countries, prevalence studies are often conducted on a national level and often relies on case-based results from individuals routinely tested for drug resistance at diagnosis. There is also a European initiative (SPREAD) to merge data and analyse on a European level since 2001 [260] and ECDC are currently investigating the possibilities of hosting a European HIV TDR surveillance network. [259, 261]

WHO has designed recommendations on TDR surveillance for LMIC, based on sampling of a limited number of individuals representing a certain population. As an increasing

proportion of individuals starting first-line ART are already ART-experienced and the information on earlier treatment is often insufficient, the updated recommendation focuses on PDR in populations initiating ART, independent of duration of infection or earlier treatment experience. [154, 253]

After introduction of ART the prevalence of TDR rose to 10-15% in Europe and USA before fully suppressive ART became available. Lately, levels in Europe have stabilized or decreased and the last figure is 8.3% 2008 - 2010, but with an increase of NNRTI resistance in recently infected patients. [260] In USA the TDR prevalence is 12 - 24% and in Eastern Europe and South/Southeast Asia 5-10%. [154]

A meta-regression-analysis of TDR surveillance studies in LMIC 2001 - 2011 showed a significant increase in prevalence of TDR, primarily NNRTI TDR, over time in SSA, starting from very low levels before ART rollout. [262] An update until 2016 included all PDR and confirmed the trend, with a predicted prevalence of NNRTI PDR for 2016 of 10.1% in eastern Africa, 11% in southern Africa and 7.2% in western and central Africa. A high NNRTI PDR prevalence of 9.4% was also estimated for Latin America and the Caribbean. NRTI PDR was stable in most areas, slowly increasing in southern and eastern Africa, but remaining <5% in all regions. PI resistance was very rare and detected in less than one percent of cases. Individuals with a reported history of ART exposure were prone to have a combination of NRTI and NNRTI resistance and were three times as likely to experience viral failure as naïve patients. [249]
WHO reports that six out of eleven LMICs surveyed for PDR worldwide 2014-2016, had NNRTI resistance at or above 10%, the critical threshold for interventions set by WHO; Argentina, Guatemala, Namibia, Nicaragua, Uganda and Zimbabwe. [263] This highlights that not only SSA but also the Americas need updated strategies on HIVDR.

The TDR prevalence in Sweden has been lower than in other parts of Europe. The overall prevalence was 5.6% in a study on data from 2003 - 2010. [264] In our study 2010 - 2016 we estimated an overall prevalence of 7.1% with a rise in NNRTI-resistance. TDR to NNRTI in migrants from SSA rose significantly from 0% in 2010 to 16% in 2016, mirroring the increase of NNRTI-resistance in SSA. [265]

A meta-analysis of more than 50,000 ART-naïve individuals from both LMIC and upper-income settings showed concordance in findings of mutations that conferred clinical resistance. M184V was the most common transmitted major NRTI-mutation and found in 54% of individuals with NRTI resistance in LMIC and 31% in high-income countries. M184I, K65R, L74V/I, Y115F and the TAMs K70R and T215Y/F were also common transmitted mutations. K103N, Y181C and G190A were the most common NNRTI-mutations in all regions and subtypes. V106M was associated to subtype C and the fourth most common NNRTI-mutation in that subtype. [246]

Generally, there is a need for standardization of methodologies for drug resistance surveillance, including continuous updates of the SDRM list and compartmentalized

analyses of specific populations. ART-naïve, undiagnosed individuals with TDR HIV are important in forwarding resistance mutations, but knowledge on the tendency of different drug resistance mutations to be maintained in transmission clusters is insufficient, especially in LMIC. [217]

1.4.8 WHO global action plan on HIV drug resistance

The rising prevalence of NNRTI resistance, especially in southern and eastern Africa but also in Latin America, threatens to undermine the efforts to achieve the 90-90-90 targets by 2020 and to end the AIDS epidemic by 2030. Mathematical modelling predicts that if no action is taken, HIVDR can result in 890,000 deaths in AIDS and 450,000 new HIV-1 infections in SSA during 2016-2030. [266] WHO has responded by a broad approach to prevent and manage HIVDR, as formulated in the WHO global action plan on HIV drug resistance 2017-2021. The objectives include strengthening laboratory capacity for access to VL and HIVDR testing as well as epidemiological surveillance and modelling to increase knowledge on HIVDR. [253]

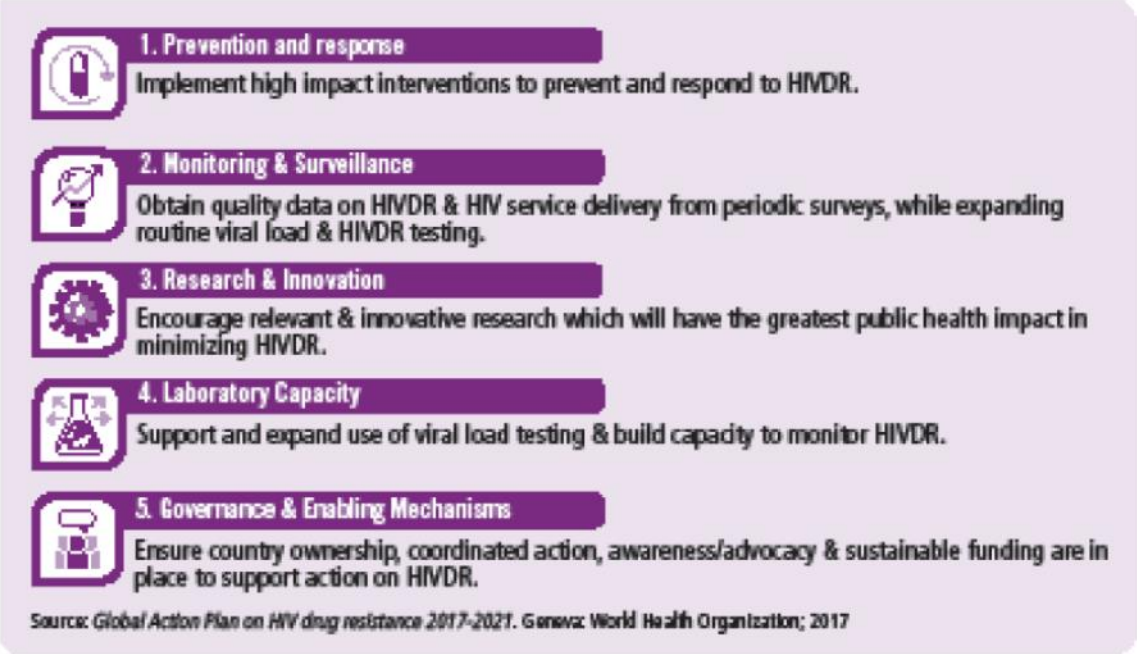

Fig 12. The five main objectives of the WHO Global Action Plan on HIV drug resistance 2017-2021. Reprinted with permission. [253]

2 AIMS

The *overall aim* of my thesis was to study molecular and epidemiological aspects of the Swedish HIV-1 epidemic in relation to the HIV-1 pandemic, using the Swedish InfCareHIV cohort. Emphasis was on pre-treatment drug resistance, undiagnosed HIV-1 infections and estimates of time to diagnosis.

The *specific* aims were:

Paper I

- To evaluate mixed basecalls in *pol* sequences as a measure of viral diversity and a method to distinguish between recent and long-term HIV-1 infection.
- To apply the method in surveys of transmitted HIV-1 drug resistance from low- and middle income countries.

Paper II

- To study transmitted drug resistance in treatment naïve individuals newly diagnosed with HIV-1 in Sweden 2010-2016.

Paper III

- To evaluate two modelling methods for estimation of the proportion of undiagnosed persons living with HIV-1 in Sweden, using national surveillance data and data from the Swedish InfCare HIV-1 cohort.

Paper IV

- To evaluate an *in-house* method for high-throughput sequencing of the HIV-1 *pol* gene with respect to clinical utility including subtype inclusivity.
- To study pre-treatment drug resistance mutations in major and minority viral variants in patients diagnosed with HIV-1 in Sweden during 2017-2019.

3 MATERIALS AND METHODS

3.1 SOURCES OF DATA

3.1.1 The national register of HIV and AIDS

Case reporting of HIV/AIDS is mandatory in Sweden since 1983 and reporting of AIDS was mandatory until 2005. National data are collected under code at the Public Health Agency of Sweden, and this register was used in **Paper III**.

3.1.2 InfCareHIV Sweden

InfCareHIV is a national decision tool, clinical database and research database initiated in 2003 by Prof Sönnerborg and with full national coverage since 2009. [71] All individuals under HIV care in Sweden are entered into the database. Historical data since the 1980's have been entered retrospectively for some clinics, e.g. in Stockholm. The database contains clinical data on diagnosis, treatment initiation and regimens, epidemiological data on origin and ethnicity of the patients, transmission mode, age, sex etc. Laboratory results such as $CD4^+$ T-cell counts, HIV-1 RNA viral loads and resistance tests are continuously entered. Sanger sequences from resistance testing of PR/RT and integrase are stored in the database. Research on the database is permitted by the steering committee after ethical approval.

InfCareHIV laid the foundation for this thesis and is the main data source for **Papers I-IV**. For these studies several downloads from the database were done to MsAccess. The downloads were done with pseudonymized patient identifiers (A five-digit code). The dates for the downloads are given in the respective papers.

3.2 LABORATORY METHODS

3.2.1 Viral load monitoring

HIV-1 viral loads are reported to InfCareHIV from several clinical laboratories that have been using different assays over the years. At Karolinska University laboratory, the NASBA® (Organon Teknika) assay was used from 1994 until 1997 when the Roche Amplicor® HIV-1 monitor test was introduced, and since then sequential versions of Cobas Amplicor® and Cobas TaqMan® (Roche Molecular Systems, Basel, Switzerland) have been used. The current assay is the Cobas®HIV-1 assay (Roche Molecular Systems, Basel, Switzerland).

3.2.2 $CD4^+$ T-cell counts

$CD4^+$ T-cell counts in InfCareHIV have been determined with routine flow cytometry. The values are for most clinics automatically transferred from the laboratory to the database.

3.2.3 HIV drug resistance testing

Population-based Sanger sequences of the *pol* gene including PR/RT were produced at Clinical Microbiology at Karolinska University Hospital, Stockholm, Sahlgrenska University

Hospital, Gothenburg or the Swedish Institute for Infectious Disease Control, Stockholm, respectively, per established *in house* protocols (1993-1999) or commercial methods (2000–2002: Trugene® HIV-1 Genotyping kit, Canada; 2003-2019: ViroSeq® HIV-1 genotyping kit, Abbott Molecular, US). Population-based IN sequences were produced with ViroSeq® HIV Integrase genotyping kit, Abbott Molecular, US. Sequences for **Paper I** were generated 1993 – 2010, sequences in **Paper II** were produced 2010-2017 and sequences for **Paper IV** 2017-2019.

In **Paper I** HIV-1 protease and reverse transcriptase sequences from 36 surveys of TDR conducted according to WHO guidelines in Africa, Asia and Latin America during 2005–2009 were included. The sequencing and base-calling of pure and mixed nucleotide positions was performed at WHO-designated laboratories from plasma or serum specimens according to local routine.

For detection of Surveillance Drug Resistance Mutations in **Paper II** the Stanford Calibrated Population Resistance Tool (https://hivdb.stanford.edu/cpr/) applying the WHO 2009 SDRM list was used. For clinical interpretations of resistance in **Papers II** and **IV**, Stanford HIVdb (https://hivdb.stanford.edu/hivdb/) was used. The current available version of the tools at the time for each study was used, for details please refer to the individual papers. For analysis of DRM in the HTS *in-house* pipeline in **paper IV**, a modified version of the list from Stanford HIVdb 8.7 was used, excluding polymorphisms in the protease of limited significance.

3.2.4 HIV-1 subtyping and phylogenetic analysis

In **Paper I** the REGA HIV-1 Subtyping Tool v2 was applied to assign subtypes. [267, 268]

In **Paper II** subtypes were determined by a maximum-likelihood phylogeny including the Los Alamos Sequence Database Subtype Reference Alignment [269], reconstructed in PhyML [270] with aLRT-SH branch support [271, 272]. Additional maximum-likelihood phylogenies were constructed for subtype/recombinant groups and cluster analysis was performed using Cluster Picker. [273] After exploratory analyses a cut-off for branch support of 0.90 and a distance threshold of 6% were chosen. [274]

In **Paper IV** The subtyping was performed using REGA version 3 and COMET. The unique recombinant forms (URFs) from HTS were confirmed using SimPlot analysis followed by fragment specific phylogenetic analysis as described by us. [27]

3.2.5 Sequence polymorphism analysis

In **Paper I**, the percentage of mixed base calls representing more than one base in a certain position (R, Y, K, M, S, W, B, D, H, V or N) in each sequence was calculated using BioEdit (http://www.mbio.ncsu.edu/BioEdit/bioedit.html) and Microsoft Excel.

3.2.6 High throughput sequencing

In **Paper IV**, a 3,000 bp fragment of *pol* including PR, RT and IN was amplified and sequenced on Illumina® HiSeq2500 producing paired end sequences of read length 250 bp. QIAamp Viral RNA extraction kit (Qiagen, Germany) was used for extraction of viral RNA. SuperScript™ III One-Step RT-PCR System with Platinum™ Taq DNA Polymerase (Life Technology,US) was used for cDNA synthesis and first round PCR with primers 1810F and 5220R. The following nested PCR applied KAPA HiFi HotStart ReadyMix PCR kit and primers 2002F and 5087R. The sequences of the primers are displayed in table 1. Purification of the amplified products was performed with QIAamp gel extraction kit (Qiagen, Germany). HTS was performed as described by our group. [275] In short, purified amplicons were fragmented and library preparation was performed using NEBNext® Ultra™ DNA Library Prep Kit for Illumina® (New England Biolab, USA) with multiplexed NEB next adaptors. During analysis, samples were pooled together with unrelated, non-viral, libraries.

Table 1. Primer information

Primer Sequence	Name	HXB2 Position
5'-GCTACACTAGAAGAAATGATGACAGCATG-3'	1810F	1810 → 1838
5'-CCCTAGTGGGATGTGTACTTCTGA-3'	5220R	5197 → 5220
5'-TGCAGGGCCCCTAGGAAAAAGGGCTGTT-3'	2002F	2002 → 2029
5'- ATCCTGTCTACYTGCCACACAAYC-3	5087R	5064 → 5087

3.2.7 Bioinformatics pipeline

In **Paper IV**, an *in-house* bioinformatics pipeline was used to process FASTQ Illumina output files. The adaptor was removed by TrimGalore version 0.6.2, and Sickle version 1.33 was used to filter low-quality bases with phred value score <Q30. FastUniq was used to remove duplicate reads. The Iterative Virus Assembler (IVA) was used for *de novo* assembly. Bowtie2 was then used to re-align the processed reads against the individual *pol* gene sequence, to create a consensus sequence covering the protease, reverse transcriptase and integrase gene (HXB2 co-ordinate: 2253- 5050). For quantification of the mutations, the selected reads were then aligned against Pol protein sequences using the blastx program from BLAST package. Best blastx hit was chosen for each read for the amino acid counting which was performed with an *in-house* script. The method provides reliable identification of viral variants down to 1% of the total viral population. [276]

3.3 MATHEMATICAL MODELS

3.3.1 CD4+ T-cell decline trajectory model

In **Paper II and IV**, a $CD4^+$ T-cell decline trajectory [74] was applied to estimate whether individuals were infected with HIV-1 before or after migration to Sweden. The method is based on the intercept and the slope of $CD4^+$ T-cell decline in seroconverters with consideration of differences depending on age and ethnicity. Individuals with primary HIV infection (PHI) or ART exposure were excluded from analysis. With input of the $CD4^+$ T-cell count at diagnosis, age and region of origin (Europe, Africa or other) the year of infection was estimated and compared to migration data.

3.3.2 SSOPHIE

A European cohort initiative, Stochastic Simulation of Outcomes of People with HIV In Europe project in EuroCoord (SSOPHIE) was formed to find improved models and resulted in development of an individual-based stochastic simulation model of the HIV-1 epidemic, based on models of HIV-progression and the effect of ART. [92] We applied this model to the Swedish HIV-1 epidemic in **Paper III**. The model is informed by multiple variables; biomarkers, treatment coverage and effects, the HIV-care continuum and risk of death from both HIV-related and non-related causes. The effects of ART and HIV pathogenesis are held fixed. Individuals of HIV-positive populations are modelled from the start of the epidemic onward. The model is run multiple times in a Bayesian framework and output is calibrated to surveillance data. It is designed for good coverage of data but can be run on limited data, which results in larger plausibility bounds. An advantage of the model is that risk groups can be modelled separately, and that the model attempts to account for migration from high-incidence countries in sub-Saharan Africa. Uncertainty in the model is displayed by a 90% plausibility range (PR) of estimates. [93]

3.3.3 ECDC HIV modelling tool

In **Paper III**, the incidence method of the ECDC HIV Modelling Tool version 1.3.0 was applied to Swedish data. The tool was developed by ECDC for modelling of European HIV-1 epidemics and is available online. (https://ecdc.europa.eu/en/publications-data/hiv-modelling-tool) It is a multi-state back-calculation model based on maximum likelihood statistics. [94, 95] The input required is the annual number of HIV diagnoses, AIDS diagnoses and concurrent HIV and AIDS diagnoses. It is recommended to also enter annual number of HIV diagnoses stratified by $CD4^+$ T-cell count, and optional, deaths in HIV-positive. The method first estimates HIV incidence over time and time to diagnosis by CD4 count strata, and then estimates the size of the HIV-positive population. Migration is not addressed in the model, but subgroups can be analyzed separately.

3.4 STATISTICAL ANALYSIS

In **Paper I**, unpaired t-test was used to compare the proportion of mixed bases in recent versus chronic HIV-1 infection. The t-test was compromised by non-normal distribution of data and hence the significance was confirmed with bootstrap analysis. The non-parametric Mann-Whitney test was used for sub-analyses of smaller groups. GraphPad online application respectively GraphPad software were used (GraphPad Software, San Diego, USA).

In **Paper II** Chi-square test was applied to investigate relationships between categorical outcomes and categorical variables. Bivariate logistic regression was used to explore associations between epidemiological and viral variables with the categorical outcome TDR. Ordered logistic regression was used to assess trends over the study period. A confidence level of 95% was chosen, with p-values less than 0.05 indicating statistical significance. Data were analyzed with StataSE 12 software (Stata Corp. College Station, USA).

In **Paper IV** Based on prior knowledge of clinical relevance and similarity to Sanger DRM testing a level of ≥20% of DRM in the viral population was chosen as primary outcome for statistical analysis of PDR. Chi-square test and unpaired t-test were applied as adequate to investigate statistical associations. When Chi-square test was compromised by too few observations, it was replaced by Fisher´s exact test. Bivariate logistic regression was used to explore associations between epidemiological and viral variables and two categorical outcomes, successful HTS and DRM in PR/RT in ≥20% of the viral population, respectively. Variables with a p-value <0.2 were included in multivariable logistic regression models, built using the "stepwise" procedure. A p-value <0.05 was considered significant in the final models. Crude and adjusted odds ratios (ORs) with 95% confidence intervals were presented.

3.5 ETHICAL CONSIDERATIONS

All four studies were conducted in accordance with the Declaration of Helsinki and approved by the Regional ethics committee in Stockholm (Diary numbers 2005/1167-31/3, 2007/1533-32, 2014/928-31, and 2017/1421-31) and by the Regional ethics committee in Gothenburg (Diary number 532-11 including amendments 20111118, 20140422, 20170516 and 20181015).

4 RESULTS AND DISCUSSION

4.1 PAPER I

"Evaluation of sequence ambiguities of the HIV-1 *pol* gene as a method to identify recent HIV-1 infection in transmitted drug resistance surveys"

Study background

To distinguish recent from long-standing HIV-1 infections is of relevance for the estimation of HIV-1 incidence, and in the study of TDR since resistance mutations may revert with time from infection. Several commonly used serological assays for identification of recent HIV-1 infection, such as the BED assay, are limited by relatively high false recent results and suboptimal performance in non-B subtypes. At the time of the study, the WHO recommendation on study of TDR in LMIC was a truncated consecutive sampling for resistace testing of ≤47 specimens from recently infected individuals within a geographic region. To ensure recent infection, epidemiological surrogate criteria were used, but there was an imminent need to ascertain their validity by an accessible biomarker. Due to the subtype diversity of the Swedish HIV-1 epidemic and research collaborations with Ethiopia and Vietnam, we had access to a suitable reference material from individuals with recent (≤1 year) and chronic (>1 year) infections.

Results and discussion

We corroborated that there was a significant difference ($p<0.0001$, unpaired t-test) in the proportion of mixed (ambiguous) base calls in routine *pol* sequences between recently ((≤1 year, n=250) and chronically (>1 year, n=441) infected individuals. This was already shown in subtype B by the Swiss Zurich Primary HIV Infection Study [113] and the Canadian HIV Strain and Drug Resistance Surveillance Program. [112]

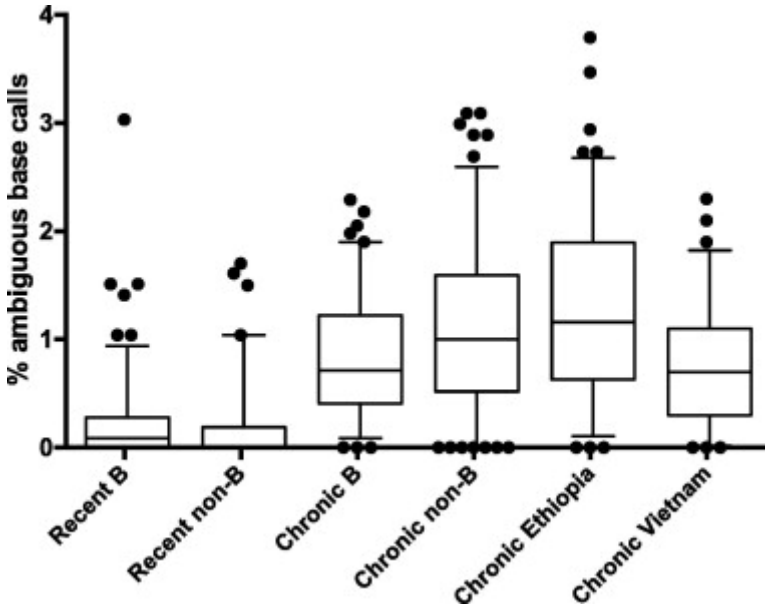

Fig 13. Proportion of mixed bases (ambiguities) in *pol* sequences from recently and chronically infected individuals. Reprinted with permission. [110]

Our main contribution was that we demonstrated that the finding was consistent over all tested non-B subtypes (A1, C and CRF01_AE), strengthening the validity of the approach in LMICs where non-B subtypes dominate. From our reference material we calculated an optimized cutoff of 0.47 % mixed bases that identified recent HIV infection with a sensitivity of 89% and a specificity of 75%.

Our analysis suggested that 71% of the individuals in TDR surveys, selected to have recent infection, were sampled within a year from infection. Interestingly, the proportion varied from 22% to 100% between surveys, suggesting that the specificity of selection criteria varied drastically between settings. Technical differences between labs may also have impacted on these regional differences.

That our reference dataset consisted of sequences from different labs and different times was a possible source of error, but also a strength since it showed the universality of the method under real-life circumstances. We chose the best possible cutoff to discriminate between recent and chronic infection in our dataset and reached a sensitivity of 89% and a specificity of 75%. This makes the method useful in an epidemiological context, especially in combination with other biomarkers, but not specific enough to use on the individual level.

The sensitivity is hampered by infections with more than one viral variant, if the infecting variants differ significantly, and a probable reason for low specificity is decreased diversity in advanced disease. Both specificity and sensitivity are negatively affected by low sequence quality and possible operator errors of over- respectively under-calling mixed bases. Since seroconversion data on chronic infections in non-B subtypes was scarce, we used $CD4^+$ T-cell counts as a surrogate variable to define this group. To ascertain that the sequences chosen were from chronically infected individuals, we chose a low $CD4^+$ count of $\leq$200 cells/mm^3, and this means that individuals with a long duration are probable to be overrepresented and we might have overestimated the difference in mixed bases between the groups.

Testing the method on a training dataset of seroconverters different from the reference dataset would have been a way to strengthen the validity. It can be argued that the dichotomized classification between recent and chronic infection is artificial and since it is proven that viral diversity increases linearly, the proportion of ambiguous base calls could also be used on a linear scale. Next generation sequencing provides opportunities to make more advanced analyses of viral diversity [114, 116] but given that population-based sequencing is still the standard method worldwide, analysis of mixed base calls is cost-effective and accessible.

4.2 PAPERS II AND IV

"Increase in transmitted drug resistance in migrants from sub-Saharan Africa diagnosed with HIV-1 in Sweden"

"High-throughput sequencing for in-depth analysis of pre-treatment HIV-1 drug resistance in Sweden."

Study background

TDR prevalence has been low in Sweden compared to other parts of the Western world, and associated to MSM and subtype B. With increasing impact of migration on the Swedish HIV-1 epidemic and reports on increasing NNRTI-resistance in SSA, we hypothesized that individuals with origin in SSA, but diagnosed with HIV-1 in Sweden would be increasingly prone to harbor NNRTI resistant virus. The nationwide coverage of the InfCareHIV database made it possible to conduct a retrospective study on naïve individuals 2010-2016 in **Paper II**.

The findings from **Paper II** and the global introduction of INSTI as a cornerstone of ART, motivated us to study resistance in both major and minority viral variants including INSTI mutations, in a well-characterized subset of newly diagnosed individuals in Sweden 2017-2019 in **Paper IV**. We embraced the WHO concept of pre-treatment drug resistance since an increasing proportion of individuals entering HIV-1 care in Sweden have an ART-history and we included both naïve and treatment experienced individuals. By applying HTS to a relevant cohort, we also aimed to evaluate our *in-house* HTS method and bioinformatics pipeline for clinical use.

Results and discussion

In **Paper II**, we showed an increase in the overall prevalence of SDRMs in naïve individuals compared to earlier data, although confidence intervals overlap. The TDR prevalence 2010-2016 was 7.1% (95% CI 5.8–8.3%) compared to 5.6% (95% CI 4.5–6.9%) in a study from 2003-2010. [264] There was a significant increase in NNRTI TDR during the study period from 1.5% (95% CI 0–2.9%) in 2010 to 6.2% (95% CI 2.9–9.5%) in 2016 (P=0.006, ordered logistic regression).

NNRTI resistance was associated with reported infection is SSA, origin in SSA and infection with subtype C. The prevalence of NNRTI TDR in individuals infected in SSA increased from no cases in 2010 to 16% (95% CI 5.0–27%) in 2016 (P=0.002, ordered logistic regression). The most common NNRTI mutation was K103N, a mutation that confers high-level resistance to EFV. Of 46 individuals from SSA with TDR, only three were reported and/or estimated with the $CD4^+$ T-cell trajectory model to have been infected after arrival to Sweden. Phylogenetic analysis did not support spread of resistant strains in Sweden within this group. Base-line resistance testing is already implemented in the Swedish HIV-1 clinics, but the limitation of population-based sequencing in detecting potentially relevant NNRTI-resistant minority variants calls for caution in using EFV in individuals with epidemiological links to SSA.

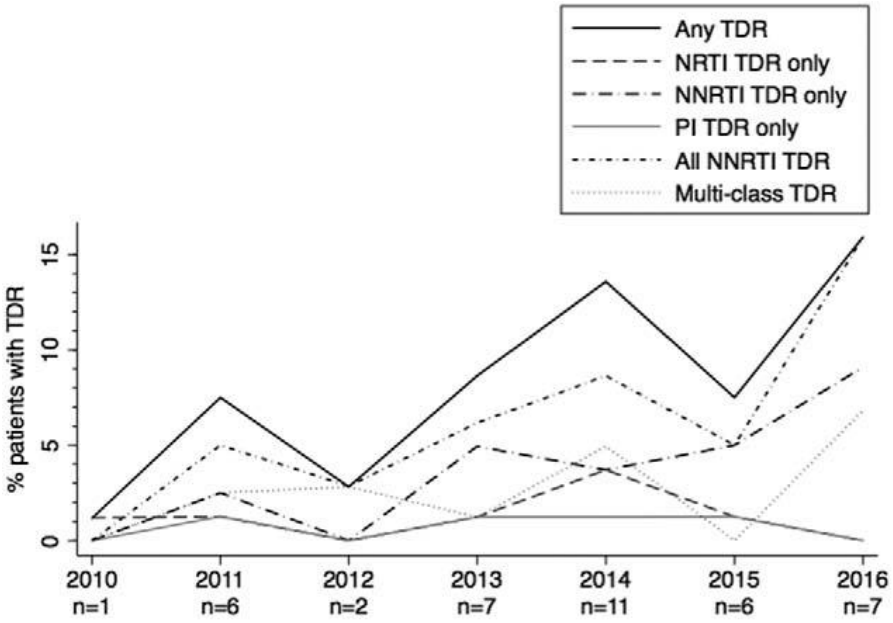

Fig 14. Prevalence of transmitted drug resistance per drug class in patients infected in sub-Saharan Africa. N=absolute number of patients with any TDR per year. TDR, transmitted drug resistance. [265] Reprinted with permission.

In contrast, NRTI resistance did not increase over the study period and was associated to infection in Sweden and subtype B. The most common singleton NRTI-mutation was M41L and a large cluster in MSM (n=26) of subtype B HIV-1 with M41L dating back to 1994 was described. M41L is a TAM and a remnant from resistance development during older ART-regimens containing thymidine analogues not in use today (i.e. AZT). M41L does not compromise modern ART, but the propagation of M41L from naïve individuals is an important proof-of-concept of the ability of fitness-neutral DRM to persist and spread clonally over decades.

In **Paper IV**, the first objective was to evaluate our *pol* HTS assay in a proper context of individuals presenting with viremic HIV-1 infection in Sweden. Our results with 87% success in producing interpretable sequences from study samples (195/224) were reassuring for subtype inclusivity. However, the method had lower success in samples with low viral loads than the currently used Sanger sequencing of PR/RT. Individuals with low viral loads were commonly on non-suppressive treatment and at high risk of HIVDR. To be able to use HTS for both clinical resistance testing and studies of resistance development in low level viremia, we need to improve the amplification step of the protocol in samples with low viral loads. It is not surprising that it is harder to amplify larger fragments (3,000bp in the HTS assay compared to 1,800bp for PR/RT and 900bp for integrase in the Sanger assay). One way to obtain HTS sequences could be to use the efficient amplification step from our Sanger protocol and then apply HTS to the amplicons. Another could be to increase input volume to increase the number of targets for PCR in the amplification step.

The second objective of **Paper IV** was to study PDR with HTS. For clarity we separated the analyses for PR/RT DRM and INSTI DRM. We found DRMs in PR/RT in 49% of the study

subjects, with all findings displayed according to proportion of the total viral population (<5%, 5-19% or ≥20%) in figure 15 a,b and c. Findings of DRM <5% were predicted by lower viral load (adjusted OR 0.61 95% CI(0.43-0.86); p=0.005) and origin in Eastern Europe or Central Asia (adjusted OR 3.72 95% CI(1.21-11.46); p=0.022). DRM 5-19% was not statistically associated to any parameter. We chose DRM ≥20% as cutoff for our analysis of PDR in relation to epidemiological and clinical data based on evidence of consensus with Sanger sequencing at this level. DRM in PR/RT in ≥20% of the viral population was found in 18% (n=36) of the study subjects. In bivariate analysis it was associated with female sex (crude OR 0.35 for men 95% CI(0.16-0.76); p=0.008), previous HIV-1 diagnosis abroad (crude OR 3.44 95% CI(1.41-8.40); p=0.007), previous ART exposure (crude OR 11.22 95% CI(3.80-33.07); p<0.0001) and to being born in Asia (crude OR 4.14 95% CI (1.13-15); p=0.032). Origin in SSA was bordering a statistically significant association (OR 2.48 (95% CI 0.96-6.45) p=0.06, bivariate logistic regression) In a multivariable logistic regression model DRM ≥20% was significantly predicted by ART exposure (adjusted OR 8.70 95% CI(2.54-29.77); p=0.001) and origin in Asia (adjusted OR 15.23 95% CI(2.13-108.67); p=0.007). Female sex and origin in SSA were associated to ART exposure and were therefore not needed for prediction and dropped from the model during the "stepwise" procedure. Individuals from Asia with PDR did not report ART exposure, suggesting that the findings in them might be due to transmitted rather than acquired resistance. PDR was in most cases due to NNRTI resistance DRMs, and the mutations and combinations of mutations found in individuals from SSA had higher predicted impact on drug effects than those found in individuals from Asia.

We wanted to study INSTI resistance and polymorphisms in a situation when few individuals are expected to have been exposed to INSTI at diagnosis in Sweden, a situation that is probable to change in the next few years due to massive roll-out of DTG in LMIC to tackle the emergence of NNRTI-resistance. We did not find any evidence of transmitted or acquired INSTI resistance, but in 8% of individuals we found major INSTI mutations in low proportions of the virus population. Of special interest is when the finding in four individuals was combined with the polymorphism L74M/I that has been suggested to contribute to failure on cabotegravir regimens in a recent study. [185] All findings in the integrase gene are displayed in figure 15 d.

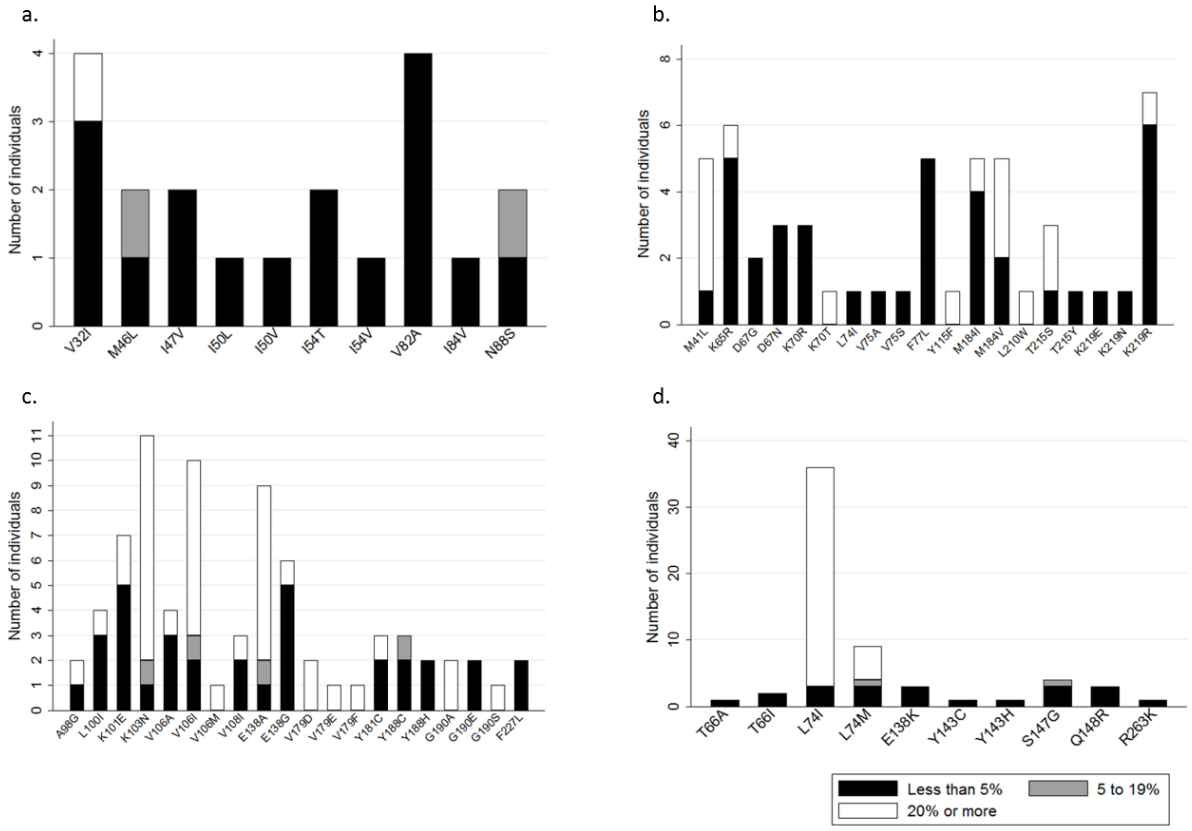

Fig 15. All findings of DRM with HTS in **Paper IV** according to their proportion of the total viral population. a) PI DRMs b) NRTI DRMs c) NNRTI DRMs d) INSTI DRMs

The roll-out of PrEP (TDF+FTC) in high risk MSM in Sweden is an important preventive intervention that was introduced during 2018. Vigilance of TDR in this group is vital, since introduction or development of resistant strains could lead to fast clonal spread of TDR HIV-1. So far there are no clinical reports of PrEP failure due to TDR in Sweden. In **Paper II** we found two MSM with strains that were highly resistant to both agents in the current PrEP regimen. In all patients, eight individuals had viruses with decreased sensitivity to TDF, eight to FTC/3TC and nine to both. In **Paper IV** one ART naïve MSM who had acquired HIV-1 in Latin America had a virus with decreased sensitivity to FTC/3TC. Two ART experienced females from SSA had FTC/3TC resistance and one heterosexual man from SSA had combined resistance to both PrEP antivirals.

Migrant MSM are of increasing importance to the Swedish HIV-1 epidemic, and in our cohort in **Paper IV,** half **(**53%) of MSM newly diagnosed with HIV-1 were of non-Swedish origin. Introduction of resistant strains from settings with emerging resistance such as SSA or Latin America as well as the increased risk of exposure to resistant HIV-1 during travel strengthen the indication of HIVDR monitoring to ensure future PrEP efficacy.

The results from **Paper IV** cannot be directly compared to the results from **Paper II**. **Paper II** aimed to study TDR (although the register-based approach is probable to have resulted in inclusion of some ART experienced individuals) over several years and to draw conclusions on general prevalence and trends of TDR in newly diagnosed individuals in Sweden. **Paper**

IV is not a true TDR prevalence study and instead aimed to study PDR in a more well-described, restricted cohort and to evaluate a new sequencing approach. However, both studies clearly show that the global emergence of NNRTI resistance impacts on the Swedish HIV-1 epidemic, and that is one of the major findings in this thesis.

4.3 PAPER III

"Challenges in modelling the proportion of undiagnosed HIV infections in Sweden."

Study background

Sweden has already reached and exceeded the second and third "90ies" of the 90-90-90 target. However, the true proportion of undiagnosed individuals i.e. the first target with 90% of individuals with HIV infection aware of their HIV status, remains elusive. The high proportion of migrants in newly diagnosed individuals in Sweden needs to be accounted for in estimating the number of undiagnosed individuals, since migrants arriving with HIV infection are only part of the Swedish undiagnosed population after arrival. We set out to estimate the undiagnosed proportion of PLHIV in Sweden with two mathematical models designed for European conditions.

Results and discussion

With the SSOPHIE model, the estimated proportion of undiagnosed PLHIV in 2013 was 26% (n=2,100; 90% PR 900-5,000) and with the ECDC HIV Modelling Tool the proportion was 21% (n=2,013; 95% CI 1,831-2,189) in 2013 and 20% (n=2,107; 95% CI 1,688-2,577) in 2016. The models are based on different statistical methods (Bayesian vs. maximum likelihood) and this explains the difference in the wide plausibility range of the SSOPHIE estimate compared to the narrow confidence interval surrounding the ECDC Modelling Tool point estimate, even though much of the input data were the same. There is no real disagreement between the methods but the estimates differ in statistical precision as measured by PR and CI, respectively.

Both models support definition of subpopulations of individuals that share epidemiological characteristics. This is an advantage in the Swedish situation with a compartmentalized epidemic. However, it is always a simplification since communication between transmission groups and misclassification do occur. In our work we decided to model MSM, heterosexual migrants from SSA and all other heterosexual transmission separately. PWID and migrants from South East Asia were too few to be modelled separately even though they are important groups in Sweden. Characteristics of individuals with heterosexual transmission that were not migrants from SSA were diverse. This group included both migrants from areas other than SSA and Swedish born, making this grouping questionable for modelling.

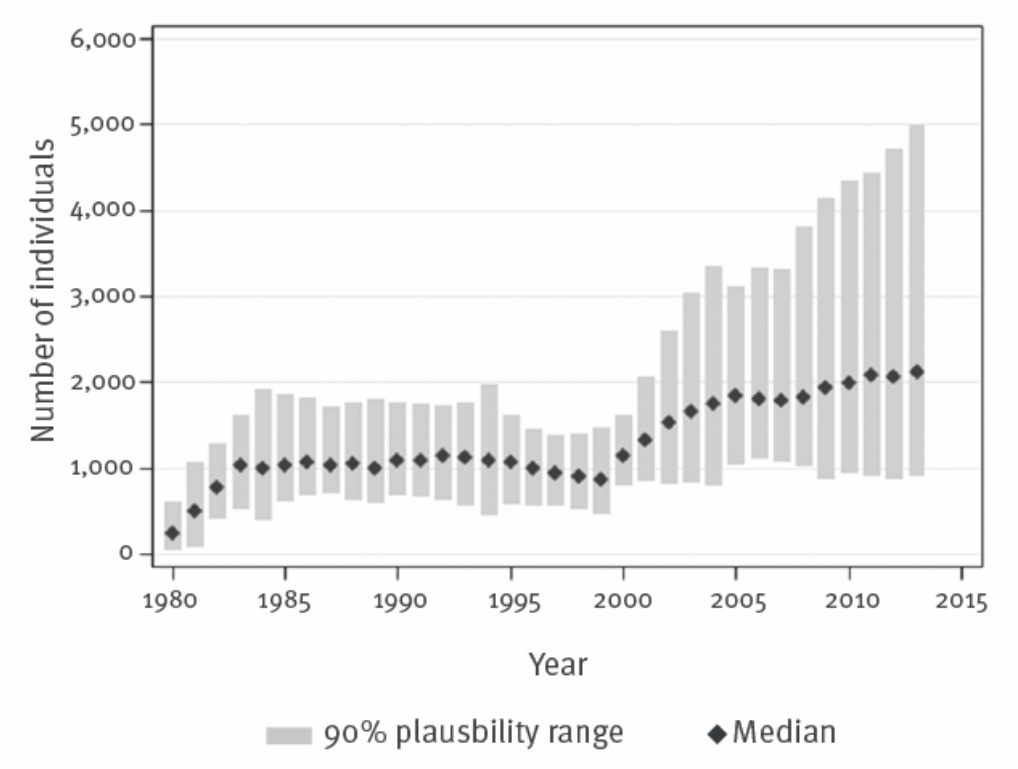

Fig 16. Number of undiagnosed PLHIV in Sweden 1980-2013 as estimated by SSOPHIE. [99] Reprinted with permission.

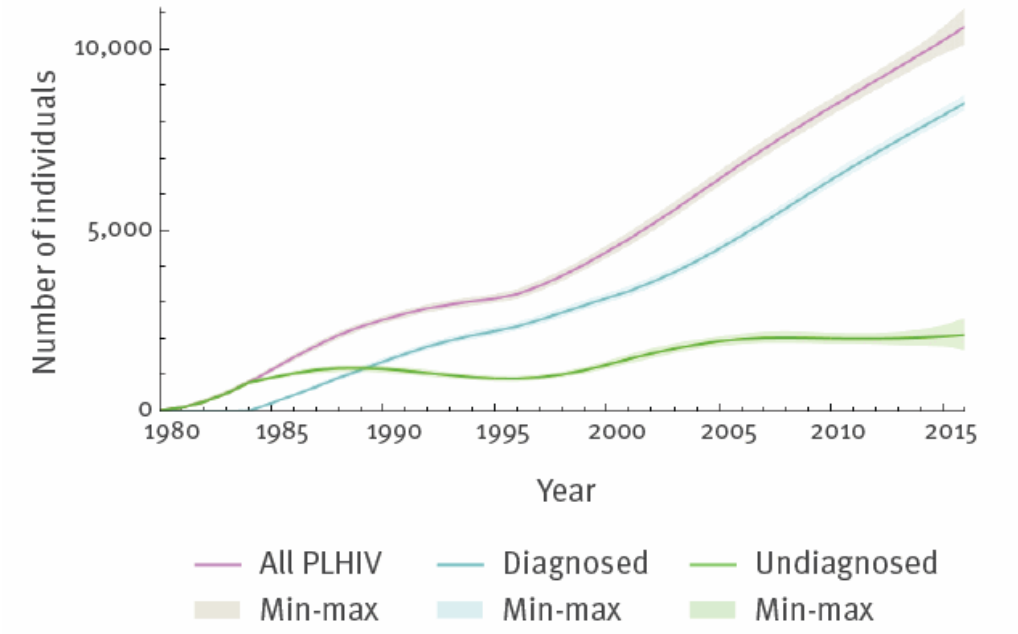

Fig 17. Number of diagnosed and undiagnosed PLHIV in Sweden 1980-2016 as estimated by ECDC HIV modelling Tool. [99] Reprinted with permission.

The results of the subpopulation modelling agreed on that the proportion of undiagnosed individuals was lowest in the MSM group; 17% (n=600; 90% PR 100-2,000) and 15% (n=369; 95% CI 299-434), in the SSOPHIE and ECDC models respectively. This is in line with our prior knowledge that MSM in Sweden with HIV-1 are less likely to present late to care than other groups of patients. [79]

The population with the highest proportion of undiagnosed infections was migrants from SSA. The SSOPHIE model produced gender specific estimates of 35% (n=300; 90% PR 200-700) in male and 34% (n=400; 90% PR (200-800) in female migrants from SSA. The ECDC modelling Tool estimated that 21% (n=530; 95% CI 436-632) of migrants from SSA with HIV-infection were undiagnosed. In this group where we anticipated an important impact of infections acquired before arrival to Sweden, the influence of migration was modelled in SSOPHIE with support of Spectrum/EPP estimates from SSA and data on migration from SSA to Sweden. This approach has several weaknesses, one being the difficulty to obtain comprehensive migration data on time of arrival to Sweden for migrants. We used data from Statistics Sweden on the number of permits to stay in Sweden issued per year to individuals from SSA as a proxy for time of arrival. However, for asylum seekers there may be a long delay from arrival to Sweden until acquisition of a permit to stay, and the delay has varied over the years. Not all migrants are granted a permit to stay after the application process and those who are declined will not be accounted for in our data. Neither did we use any data on emigration from Sweden.

The assumptions regarding migration are vital to produce correct estimates on the size of the undiagnosed group. As is increasingly emphasized, the HIV epidemics in countries with generalized epidemics are also diverse and prevalence may differ greatly between geographical areas and populations within one country, meaning that using general prevalence in the country of origin for migrants may be misleading.

To our disappointment, we did not obtain an estimate of the undiagnosed proportion of HIV in Sweden that was reliable enough to use for public health purposes. The uncertainty bounds of the SSOPHIE estimate were wide, and the ECDC-model did not take migration into account. Compared to our preconceptions the estimated proportions of undiagnosed PLHIV from the models were unexpectedly high.

The way forward towards more reliable estimates lies in development of both new models and appropriate collection of surveillance data to meet the challenges of producing estimates in a country with small numbers of PLHIV and massive impact of migration from all over the world.

5 CONCLUSIONS

- Viral diversity measured by the proportion of mixed bases in routinely generated population-based *pol* sequences discriminates reasonably well between recent and chronic HIV-1 infection in the major subtypes. Further improvements might include:
  - Combinations with other biomarkers that could strengthen the estimates.
  - Standardization of sequencing to minimize technical error.
  - Viral diversity assessed with NGS/HTS is a promising development of the approach.

- Transmitted drug resistance to NNRTI in migrants from sub-Saharan Africa increased substantially between 2010 – 2016.
  - The global trend in NNRTI resistance is reflected in migrants in Sweden.
  - Sweden has potential to be a beacon for future resistance trends.

- Pre-treatment drug resistance can be analysed in a wide range of HIV-1 subtypes with our *in-house* NGS/HTS assay.
  - Pre-treatment drug resistance in PR/RT in our cohort was predicted by ART exposure and origin in Asia.
  - NNRTI resistance is the most clinically important finding and occurs in individuals from all over the world.
  - No evidence of transmitted or acquired resistance to INSTI were found, and few individuals were exposed to INSTI.

- Cases of HIV-1 with resistance to one or both components of PrEP are found in newly diagnosed MSM in Sweden.
  - Migration and travel are important in shaping the Swedish HIV-1 epidemic in MSM.
  - Vigilance of TDR affecting PrEP efficacy in the MSM group is important.

- The proportion of undiagnosed people living with HIV in Sweden is still uncertain.
  - Mathematical models suitable for low-prevalence settings with high impact of migration need further development.
  - High quality case-based data on migration in HIV-positive migrants as well as population-based data on general migration from high-endemic countries is needed for reliable estimates.

6 FUTURE PERSPECTIVES

During the work with this thesis important changes in HIV-1 epidemiology and drug resistance epidemiology and a rapid technical development have taken place both in Sweden and globally and these trends are of importance for future research, treatment and prevention of HIV-1 in Sweden.

Migration and travel are central components of the HIV-1 pandemic and Sweden has the highest proportion of migrants among newly diagnosed PLHIV in Europe. An increasing proportion of migrants initiating HIV care in Sweden are already diagnosed and many are on ART, but information about this is insufficient. The classical epidemiological compartmentalization in Swedish MSM vs heterosexual migrants from high-endemic settings is challenged by an increasing proportion of migrant MSM in the population with new HIV-1 diagnoses in Sweden. In both migrants and Swedish born with new HIV-1 diagnoses it is often uncertain if the HIV-1 infection was contracted in Sweden or abroad. Better methods to determine when an individual was infected in relation to migration and travel is still a high priority. This important question is addressed by our group as a main objective in the ongoing prospective TIME-study, where time to diagnosis in newly diagnosed PLHIV is estimated with a combination of NGS viral diversity, serological markers and $CD4^+$ T-cell counts in a Bayesian framework. We hope that this will create a robust method that can be integrated in epidemiological studies of HIV-1 as well as in surveillance activities in Sweden. In addition, more detailed collection of surveillance data of HIV-1 in Sweden to identify previously diagnosed infections and individuals already under effective would increase possibilities to estimate undiagnosed infections and to design effective interventions for early diagnosis and prevention.

NGS/HTS is rapidly moving from an exclusive research technology into an everyday clinical tool. We intend to use the experiences from **Paper IV** to introduce NGS/HTS resistance testing of HIV-1 in the clinic within the next few years. However, our results with limited success in samples with low viral loads show that improvements are needed to fulfill clinical requirements. Routine collection of NGS/HTS data would facilitate surveillance of PDR and make it possible to detect a possible increase in PDR to integrase inhibitors or other important drugs before it has large clinical consequences. Collection of more data will help to further elucidate the clinical significance of low level DRMs and the adequate level of reporting. It is probable that the clinical importance of a certain percentage of resistant variants in a position is related not only to the tested drug, but to potency of companion drugs, viral load, adherence, and potentially other host factors and viral subtype. Machine learning approaches is an interesting approach to find patterns in big data sets and could potentially be used to investigate this.

PrEP is so far a success story in preventing HIV-1 transmission in MSM and has been well-received when recently introduced in Sweden. It has the potential to prevent many of the transmissions among MSM residing in Sweden, both in-country transmissions and during

travel. To ensure that the PrEP regimens are effective, up-to-date surveillance of TDR is a necessity and this could be accomplished by the methods we have used in this thesis.

Baseline resistance testing is a cornerstone of HIVDR management in a high-income setting, but an increasing proportion of HIV positive individuals presenting for care in Sweden are already diagnosed and on suppressive ART. This is good news but means that baseline resistance testing and complete treatment history is often not available. In an era of treatment simplification, proviral resistance testing is an interesting concept, but not without caveats, and needs to be further investigated for clinical use.

Time is closing in on the global achievements of the 90-90-90 target goals. It is highly unlikely that they will be reached universally by next year. Sweden is in a good position with excellent treatment rates and viral suppression rates. However, many individuals are still being diagnosed late in Sweden and the number and characteristics of undiagnosed people living with HIV is unclear. Joint efforts in epidemiology, diagnostic and educational activities are needed to enumerate and diminish undiagnosed infections. The global progress towards 95-95-95 in 2030 depends on the success in tackling HIVDR and on elimination of stigma and political obstacles that prevent access to high-quality HIV services for all PLHIV.

The diverse population of PLHIV in Sweden, combined with a strong national surveillance program and a national database collecting real-time clinical data, provides opportunities for future research of global interest although the number of PLHIV in Sweden is relatively small.

7 ACKNOWLEDGEMENTS

The writing of this thesis and the completion of my PhD studies would not have been possible without blood (EDTA), sweat and even occasional tears, but more importantly the professional and moral support from a lot of people.

To start with, my perfect Dream Team of supervisors:
Anders Sönnerborg, my main supervisor, you are a role model in translational HIV research, always with the patient in focus. You have given me plenty of guidance and challenged me to do things I didn´t know I could. You really made me grow as a professional and as a person, through your early advice "don´t be shy". I´ve always trusted your good judgement and your sense of humor is a great antidote to despair in difficult times.
Jan Albert, co-supervisor, you are a true inspiration in how to conduct research in clinical virology. Your contagious enthusiasm and technical knowledge paired with your patience in supervision have made invaluable contributions to my professional development. Your writing and presentation skills also had massive impact on me. To only have one slide per two minutes remains the ultimate (impossible?) goal for future presentations.
Johanna Brännström, co-supervisor, besides contributing high-class epidemiological skills and an uncompromising stand on data quality, you´ve been the best moral support one could wish for. I couldn´t think of a better partner in carrying out a clinical epidemiological study with all its challenging aspects. I also admire your true activist soul and count on continued collaborations.

All people living with HIV in Sweden; your generous contribution of data to InfCareHIV laid the foundation for this work. Special thanks to the participants in the TIME study.

Some of the other people to whom I am thankful are:
My official mentor in my PhD studies, ST-supervisor in Clinical Virology and long-term roommate **Agneta Samuelson**, your experience and support in all aspects of diagnostics, research and life have been invaluable. I admire your unprecedented stringency in attacking any diagnostic issue and wish to once be as well-organized in my thoughts.
My boss at Karolinska University Laboratory **Ilona Lewensohn-Fuchs** for helping me pursue my specialist training in both Clinical Microbiology and Infectious Diseases, while sending subliminal signals that made me end up in Clinical virology after all. Your flexibility and pragmatism as well as your outspoken support for research made this thesis possible.

Michael R Jordan, Silvia Bertagnolio and the rest of the WHO Drug Resistance Team, for welcoming me as an intern in Geneva in 2010. Together with the rest of the team that contributed to the first paper; **Wei Shao, Irene Bontell, Fatim Cham, Do Duy Cuong, Amogne Wondwossen, Lynn Morris, Gillian Hunt and Frank Maldarelli,** you helped me get started with research and made me find the right angle for this thesis.
Joakim Esbjörnsson for inviting me to Lund for an intense individual crash-course in phylogeny, and for amazing distance supervision. **Agnes Nordquist, Leo Flamholc,**

Magnus Gisslén, Bo Hejdeman, Gaetano Marrone, Hans Norrgren, Veronica Svedhem, Suzanne Wendahl for invaluable contributions to the second paper.
Fumiyo Nakagawa and **Andrew Phillips** for performing the modelling of our data in the third paper and for patience in explaining Bayesian epidemiological modelling. **Ard van Sighem** for all the help with the ECDC HIV Modelling Tool.

All past and present members of the dynamic Sönnerborg research group, especially **Ujjwal Neogi**, **Shambhu Aralaguppe** and **Anoop Ambikan** for fruitful collaborations and intense work on the last manuscript. **Amanda Häggblom**, for good advice in research and many supportive lunches with discussions on HIV and all other aspects of life. **Tina Carlander** for being a cool role model in research. **Kajsa Noyan** for nice chats and support.
The Albert research team; **Lina Thebo, Tanja Normark** and **Eva Eriksson** for all your efforts with laboratory and bioinformatics analyses of the TIME samples.

All my colleagues in Clinical microbiology Huddinge and Solna, for your support and enthusiasm in the daily work. Special thanks to the virology crew at F68; **Zhibing Yun**, **Anna Karlsson**, **Agnes Bölin-Wiener**, **Niklas Björkström**, **Gustaf Sandh, Martin Eklund, Johan Aarum, Chammiran Daniel and Harald Eberhardt**. The heroes of the HIV/hepatitis diagnostic section, especially **Ann Gimblad**, **Anne Quist**, **Sandra Johansson**, **Annie Hsieh**, **Linnea Törner Larsson**, **Claire Keane, Nikolina Derikonja** and **Sozanah Rashid** for fruitful discussions that developed my diagnostic skills - and for caring for "my" precious TIME-study samples. **Peter Bergman** for inspiration and support. **Eileen Belardo** for solving everything, every time. **Robert Dyrdak** for sharing my enthusiasm in virology, NGS and database design and for diagnosing me with Z73.2 rather than Z73.4. **Volkan Özenci** for trusting me with student education and for general support and encouragement.

Everyone at the Infectious Disease department of Karolinska University Hospital, for all the experiences and skills that I´ve acquired over the years. Special thanks to superhero **Malin Ackefors**, supervising my ID specialization. **Jenny Stenkvist**, former ID colleague and present Infection Control guru, for your enthusiasm around viruses initialized with H, and your interest in undiagnosed infections.
The TIME-study would not have been possible without all the hard-working HIV doctors at Karolinska, Venhälsan and Sahlgrenska, nor without the dedicated study nurses. Thank you all, especially **Veronica Svedhem, Piotr Nowak, Jan Vesterbacka , Anders Thalme**, **Lena Mattsson** and **Maria Thorell Amsenius** at Karolinska, **Aylin Yilmaz** in Gothenburg, **Göran Bratt, Finn Filén, Catharina Missailidis, Elisabeth Storgärd, Bernt Hildingsson-Lundh** at Venhälsan.

Arja Kramsu, **Andrej Weintraub and Marita Wallenberg Lundgren** at the department of Laboratory Medicine, for making the double-life of the clinical PhD-student less complicated.

Gunilla Källenius for introducing me to molecular epidemiology once upon a time at Smittskyddsinstitutet – I got stuck, I only changed the bug!
Maria Axelsson and **Torsten Berglund** at the Public Health Agency of Sweden for all your guidance regarding undiagnosed proportions and for digging deep into the original records from the 1980´s as well as providing smoking hot surveillance data whenever needed.

All colleagues in ESCMID (European Society of Clinical Microbiology and Infectious Diseases) especially **Gunnar Kahlmeter, Murat Akova, Jesus Rodriguez-Bano** and **Mario Poljak**, for welcoming me into the society and trusting me with all kinds of challenging tasks that contributed to my evolution into a clinical virologist.

The **KI/SLL Research school in Molecular biology for Clinicians** for giving me the opportunity to broaden the view and for providing both useful research tools and many new friends and contacts. And for the Zebra fish experience!

My wise and intellectual friends in "**Bokcirkel för Annas vänner"**: **Anna Svensson, Anna Lindam, Eva-Lena Engman** and **Jenny Andersson**, for being a constant through all stages of life – there is always a book and a café and a lot to talk about (sometimes related to the book we´ve (maybe) read)!
My friends from the very beginning: **Julia Englund** and **Mia Helenius Lockner** for support and inspiration even though time has been scarce lately – let´s hang out more!
Linda Eriksson for longstanding friendship, encouragement and companionship since the very first day in medschool. Your integrity and grit is always an inspiration and I am looking forward to your thesis defence soon!
Friends **Rahman** and **Ebba Amandius** for inspiration in science and interesting discussions throughout the years. Thank you for good advice on the summary in Swedish!
Inga Fröding and **Caroline Rönnberg**; for better or worse through the whole PhD process, your support and encouragement has been outstanding. Let´s go to that funny-hat party at the town hall and drink lots of champagne!

My parents **Ulla Stina** and **Jan-Åke** for always being more interested in who I am than what I do, and for early introduction of the habit of looking everything up in the encyclopedia.
My brother **Klas** for being the best brother and friend in all stages of life. His husband **Mats** for lightening up our lives at every visit.
My parents in law **Gunnel** and **Sven-Olof** for frequently saving the day, including providing food for the whole family.
Johan, for 24/7 love, support and IT-support. You´re the one!
Tove and **Liv** for being wonderful in general, distracting me from work and reminding me that curiosity is the main driving force in science.

8 REFERENCES

1. UNAIDS. *Global HIV & AIDS statistics — 2019 fact sheet*. 2019 [cited 2019 August 20].
2. UNAIDS, *UNAIDS data 2018*. 2018: http://www.unaids.org/sites/default/files/media_asset/unaids-data-2018_en.pdf.
3. Bavinton, B.R., et al., *Viral suppression and HIV transmission in serodiscordant male couples: an international, prospective, observational, cohort study.* Lancet HIV, 2018. **5**(8): p. e438-e447.
4. Rodger, A.J., 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, 2016. **316**(2): p. 171-81.
5. Rodger, A.J., 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. **393**(10189): p. 2428-2438.
6. Murray, P.B., E; Jorgensen, J; Pfaller, M; Yolken, R, *Manual of Clinical Microbiology, 8th edition*. Vol. 2. 2003: ASM Press.
7. Richman, D., R. Whitley, and F. Hayden, *Clinical Virology, 3rd edition*. 2009. ASM Press.
8. Sharp, P.M. and B.H. Hahn, *Origins of HIV and the AIDS pandemic.* Cold Spring Harb Perspect Med, 2011. **1**(1): p. a006841.
9. Worobey, M., et al., *Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960.* Nature, 2008. **455**(7213): p. 661-4.
10. Keele, B.F., et al., *Chimpanzee reservoirs of pandemic and nonpandemic HIV-1*. Science, 2006. **313**(5786): p. 523-6.
11. Pineda-Pena, A.C., et al., *On the contribution of Angola to the initial spread of HIV-1*. Infect Genet Evol, 2016. **46**: p. 219-222.
12. Faria, N.R., et al., *HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations.* Science, 2014. **346**(6205): p. 56-61.
13. Pepin, J., et al., *Iatrogenic transmission of human T cell lymphotropic virus type 1 and hepatitis C virus through parenteral treatment and chemoprophylaxis of sleeping sickness in colonial Equatorial Africa.* Clin Infect Dis, 2010. **51**(7): p. 777-84.
14. de Sousa, J.D., et al., *High GUD incidence in the early 20 century created a particularly permissive time window for the origin and initial spread of epidemic HIV strains.* PLoS One, 2010. **5**(4): p. e9936.
15. Gilbert, M.T., et al., *The emergence of HIV/AIDS in the Americas and beyond.* Proc Natl Acad Sci U S A, 2007. **104**(47): p. 18566-70.
16. Churchill, M.J., et al., *HIV reservoirs: what, where and how to target them.* Nature Reviews Microbiology, 2015. **14**: p. 55.
17. Chan, D.C. and P.S. Kim, *HIV Entry and Its Inhibition.* Cell, 1998. **93**(5): p. 681-684.
18. Wyatt, R. and J. Sodroski, *The HIV-1 Envelope Glycoproteins: Fusogens, Antigens, and Immunogens.* Science, 1998. **280**(5371): p. 1884-1888.
19. Arif, M.S., et al., *Pace of Coreceptor Tropism Switch in HIV-1-Infected Individuals after Recent Infection.* Journal of Virology, 2017. **91**(19).
20. Finzi, D., et al., *Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy.* Science, 1997. **278**(5341): p. 1295-300.
21. Zanini, F., et al., *In vivo mutation rates and the landscape of fitness costs of HIV-1.* Virus Evol, 2017. **3**(1): p. vex003.
22. Cuevas, J.M., et al., *Extremely High Mutation Rate of HIV-1 In Vivo.* PLoS Biol, 2015. **13**(9): p. e1002251.

23. Smyth, R.P., M.P. Davenport, and J. Mak, *The origin of genetic diversity in HIV-1.* Virus Res, 2012. **169**(2): p. 415-29.

24. Carr, J.K., et al., *Reference Sequences Representing the Principal Genetic Diversity of HIV-1 in the Pandemic.* 1998, Los Alamos HIV Sequence database: https://www.hiv.lanl.gov/content/sequence/HIV/REVIEWS/SUBTYPE_98/subtypes.html.

25. Robertson, D.L., et al., *HIV-1 nomenclature proposal.* Science, 2000. **288**(5463): p. 55-6.

26. Hemelaar, J., et al., *Global trends in molecular epidemiology of HIV-1 during 2000-2007.* Aids, 2011. **25**(5): p. 679-89.

27. Neogi, U., et al., *Recent increased identification and transmission of HIV-1 unique recombinant forms in Sweden.* Sci Rep, 2017. **7**(1): p. 6371.

28. Foley, B., et al., *HIV Sequence Compendium 2018.* 2018, Los Alamos National, Laboratory, Theoretical Biology and Biophysics, Los Alamos, New Mexico.

29. Hemelaar, J., et al., *Global and regional molecular epidemiology of HIV-1, 1990-2015: a systematic review, global survey, and trend analysis.* Lancet Infect Dis, 2019. **19**(2): p. 143-155.

30. Ayehunie, S., et al., *New subtype of HIV-1 in Ethiopia.* Lancet, 1990. **336**(8720): p. 942.

31. Kalu, A.W., et al., *Monophylogenetic HIV-1C epidemic in Ethiopia is dominated by CCR5-tropic viruses-an analysis of a prospective country-wide cohort.* BMC Infect Dis, 2017. **17**(1): p. 37.

32. Lebedev, A., et al., *Human Immunodeficiency Virus-1 Diversity in the Moscow Region, Russia: Phylodynamics of the Most Common Subtypes.* Front Microbiol, 2019. **10**: p. 320.

33. Beloukas, A., et al., *Molecular epidemiology of HIV-1 infection in Europe: An overview.* Infect Genet Evol, 2016. **46**: p. 180-189.

34. Esbjornsson, J., et al., *HIV-1 transmission between MSM and heterosexuals, and increasing proportions of circulating recombinant forms in the Nordic Countries.* Virus Evol, 2016. **2**(1): p. vew010.

35. Smyth, R.P. and M. Negroni, *A step forward understanding HIV-1 diversity.* Retrovirology, 2016. **13**: p. 27.

36. Domingo, E., J. Sheldon, and C. Perales, *Viral quasispecies evolution.* Microbiol Mol Biol Rev, 2012. **76**(2): p. 159-216.

37. Karlsson, A.C., et al., *Characterization of the viral population during primary HIV-1 infection.* Aids, 1998. **12**(8): p. 839-47.

38. Keele, B.F., et al., *Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection.* Proc Natl Acad Sci U S A, 2008. **105**(21): p. 7552-7.

39. Salazar-Gonzalez, J.F., et al., *Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing.* J Virol, 2008. **82**(8): p. 3952-70.

40. Abrahams, M.R., et al., *Quantitating the multiplicity of infection with human immunodeficiency virus type 1 subtype C reveals a non-poisson distribution of transmitted variants.* J Virol, 2009. **83**(8): p. 3556-67.

41. Klein, K., et al., *Higher sequence diversity in the vaginal tract than in blood at early HIV-1 infection.* PLoS Pathog, 2018. **14**(1): p. e1006754.

42. Piantadosi, A., et al., *Metagenomic sequencing of HIV-1 in the blood and female genital tract reveals little quasispecies diversity during acute infection.* J Virol, 2018.

43. Carlson, J.M., et al., *HIV transmission. Selection bias at the heterosexual HIV-1 transmission bottleneck.* Science, 2014. **345**(6193): p. 1254031.

44. Shankarappa, R., et al., *Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection.* J Virol, 1999. **73**(12): p. 10489-502.

45. Zanini, F., et al., *Population genomics of intrapatient HIV-1 evolution.* Elife, 2015. **4**.

46. Joos, B., et al., *HIV rebounds from latently infected cells, rather than from continuing low-level replication.* Proc Natl Acad Sci U S A, 2008. **105**(43): p. 16725-30.

47. Karlsson, A.C., et al., *Reappearance of founder virus sequence in human immunodeficiency virus type 1-infected patients.* J Virol, 1999. **73**(7): p. 6191-6.

48. Magiorkinis, G., et al., *The global spread of HIV-1 subtype B epidemic.* Infect Genet Evol, 2016. **46**: p. 169-179.

49. Angelis, K., et al., *Global Dispersal Pattern of HIV Type 1 Subtype CRF01_AE: A Genetic Trace of Human Mobility Related to Heterosexual Sexual Activities Centralized in Southeast Asia.* J Infect Dis, 2015. **211**(11): p. 1735-44.

50. World Health Organization webpage; www.who.int/hiv/data/en/ [cited 2019 August 20].

51. World Health Organization Regional Office for Africa, Brazzaville, *Report on key populations in African HIV/AIDS : national strategic plans, July 2018.* 2018. Licence: CC BY-NC-SA 3.0 IGO.

52. UNAIDS. *UNAIDS AIDSinfo*. Webpage. [cited 2019 August 20].

53. Dwyer-Lindgren, L., et al., *Mapping HIV prevalence in sub-Saharan Africa between 2000 and 2017.* Nature, 2019. **570**(7760): p. 189-193.

54. Cuadros, DF., *Assessing the role of geographical HIV hot-spots in the spread of the epidemic, CROI 2018*. 2018: Boston (MA), USA. Conference abstract.

55. Samuel, R., et al., *Minority HIV-1 drug-resistant mutations and prevention of mother-to-child transmission: perspectives for resource-limited countries.* AIDS Rev, 2014. **16**(4): p. 187-98.

56. Kanthula, R., et al., *Persistence of HIV drug resistance among South African children given nevirapine to prevent mother-to-child-transmission.* Aids, 2017. **31**(8): p. 1143-1148.

57. Paraskevis, D., et al., *Tracing the HIV-1 subtype B mobility in Europe: a phylogeographic approach.* Retrovirology, 2009. **6**(1): p. 49.

58. Thomson, M.M., et al., *New insights into the origin of the HIV type 1 subtype A epidemic in former Soviet Union's countries derived from sequence analyses of preepidemically transmitted viruses.* AIDS Res Hum Retroviruses, 2007. **23**(12): p. 1599-604.

59. McCutchan, F.E., et al., *Diversity of the envelope glycoprotein among human immunodeficiency virus type 1 isolates of clade E from Asia and Africa.* J Virol, 1996. **70**(6): p. 3331-8.

60. National Health Commission of the People´s Republic of China, *Routine press conference on 23 November 2018 and Introduce the progress of AIDS prevention and control in China. In Chinese.* 2018: http://www.moh.gov.cn/xwzb/webcontroller.do?titleSeq=11139&gecstype=1. Press Release.

61. UNAIDS, *UNAIDS reports that reaching Fast-Track Targets will avert nearly 28 million new HIV infections and end the AIDS epidemic as a global health threat by 2030*. 2014: http://www.unaids.org/en/resources/presscentre/pressreleaseandstatementarchive/2014/november/20141118_PR_WAD2014report.

62. UNAIDS, *90-90-90 An ambitious treatment target to help end the AIDS epidemic*. 2014: http://www.unaids.org/en/resources/909090.

63. Lazarus, J.V., et al., *Beyond viral suppression of HIV - the new quality of life frontier.* BMC Med, 2016. **14**(1): p. 94.

64. UNAIDS, *Understanding FastTrack*. 2015: https://www.unaids.org/sites/default/files/media_asset/201506_JC2743_Understanding_FastTrack_en.pdf.

65. Giordano, T.P., M.E. Suarez-Almazor, and R.M. Grimes, *The population effectiveness of highly active antiretroviral therapy: Are good drugs good enough?* Current HIV/AIDS Reports, 2005. **2**(4): p. 177-183.

66. Ulett, K.B., et al., *The therapeutic implications of timely linkage and early retention in HIV care.* AIDS Patient Care STDS, 2009. **23**(1): p. 41-9.

67. World Health Organization, *Cascade data use manual: to identify gaps in HIV and health services for programme improvement*. 2018: http://www.who.int/hiv/pub/toolkits/hiv-cascade-data-use-manual/en/.

68. Gisslen, M., et al., *Sweden, the first country to achieve the Joint United Nations Programme on HIV/AIDS (UNAIDS)/World Health Organization (WHO) 90-90-90 continuum of HIV care targets.* HIV Med, 2017. **18**(4): p. 305-307.

69. Brown, A.E., et al., *HIV in Europe and Central Asia: progress in 2018 towards meeting the UNAIDS 90-90-90 targets.* Euro Surveill, 2018. **23**(48).

70. Pulic Health Agency of Sweden. *HIV statistics Public Health Agency of Sweden*. 2019 [cited 2019 June 26]. In Swedish.

71. Haggblom, A., et al., *HIV drug therapy duration; a Swedish real world nationwide cohort study on InfCareHIV 2009-2014.* PLoS One, 2017. **12**(2): p. e0171227.

72. Von Krogh, G., et al., *The introduction of HIV during 1979-80 in a sexually active homosexual population of Stockholm.* Scand J Infect Dis, 1987. **19**(3): p. 285-8.

73. Giesecke, J., *Den svenska HIV-epidemin I backspegeln: Små "kärngrupper" står för den inhemska spridningen.* Läkartidningen, 1998. **95**(15). In Swedish.

74. Brannstrom, J., et al., *A high rate of HIV-1 acquisition post immigration among migrants in Sweden determined by a CD4 T-cell decline trajectory model.* HIV Med, 2017.

75. Alvarez-Del Arco, D., et al., *High levels of postmigration HIV acquisition within nine European countries.* Aids, 2017. **31**(14): p. 1979-1988.

76. Fakoya, I., et al., *A systematic review of post-migration acquisition of HIV among migrants from countries with generalised HIV epidemics living in Europe: mplications for effectively managing HIV prevention programmes and policy.* BMC Public Health, 2015. **15**: p. 561.

77. Kaberg, M., et al., *Incidence and spontaneous clearance of hepatitis C virus (HCV) in people who inject drugs at the Stockholm Needle Exchange-Importance for HCV elimination.* J Viral Hepat, 2018. **25**(12): p. 1452-1461.

78. *Swedish Communicable Diseases Act*. 2004: http://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/smittskyddslag-2004168_sfs-2004-168. In Swedish.

79. Brannstrom, J., et al., *Deficiencies in the health care system contribute to a high rate of late HIV diagnosis in Sweden.* HIV Med, 2016. **17**(6): p. 425-35.

80. Neogi, U., et al., *Temporal trends in the Swedish HIV-1 epidemic: increase in non-B subtypes and recombinant forms over three decades.* PLoS One, 2014. **9**(6): p. e99390.

81. Gail, M.H. and R. Brookmeyer, *Methods for projecting course of acquired immunodeficiency syndrome epidemic.* J Natl Cancer Inst, 1988. **80**(12): p. 900-11.

82. Gail, M.H., P.S. Rosenberg, and J.J. Goedert, *Therapy may explain recent deficits in AIDS incidence.* J Acquir Immune Defic Syndr, 1990. **3**(4): p. 296-306.

83. Downs, A.M., et al., *Reconstruction and prediction of the HIV/AIDS epidemic among adults in the European Union and in the low prevalence countries of central and eastern Europe.* Aids, 1997. **11**(5): p. 649-62.

84. Taffe, P. and M. May, *A joint back calculation model for the imputation of the date of HIV infection in a prevalent cohort.* Stat Med, 2008. **27**(23): p. 4835-53.

85. Hall, H.I., et al., *Estimation of HIV incidence in the United States.* Jama, 2008. **300**(5): p. 520-9.

86. Phillips, A.N., et al., *HIV in the UK 1980-2006: reconstruction using a model of HIV infection and the effect of antiretroviral therapy.* HIV Med, 2007. **8**(8): p. 536-46.

87. Nakagawa, F., *Estimation of the size and characteristics of HIV-positive populations in Europe*, in *Research Department of Infection and Population Health*. 2015, University College London.

88. UNAIDS, *UNAIDS quarterly update on HIV epidemiology / 1Q 2010*. 2010: http://www.unaids.org/sites/default/files/media_asset/epi_alert_1stqtr2010_en_0.pdf.

89. Stover, J., *Projecting the demographic consequences of adult HIV prevalence trends: the Spectrum Projection Package.* Sexually Transmitted Infections, 2004. **80**(suppl 1): p. i14-i18.

90. Stover, J., et al., *Updates to the Spectrum/Estimations and Projections Package model for estimating trends and current values for key HIV indicators.* Aids, 2017. **31 Suppl 1**: p. S5-s11.

91. Archer, C., et al., *AIM. A Spectrum module for creating HIV/AIDS projections and examining the demographic and social impacts of AIDS.* . Futures Institute, Bethesda (MD): www.avenirhealth.org/Download/Spectrum/Manuals/AIMManualEnglish.pdf.

92. Nakagawa, F., et al., *A Method to Estimate the Size and Characteristics of HIV-positive Populations Using an Individual-based Stochastic Simulation Model.* Epidemiology, 2016. **27**(2): p. 247-56.

93. Nakagawa, F., *An epidemiological modelling study to estimate the composition of HIV-positive populations including migrants from endemic settings.* Aids, 2017. **31**(3): p. 417-425.

94. *ECDC HIV modelling tool [software application]. Version 1.3.0 Stockholm: European Centre for Disease Prevention and Control; 2017.* . 2017: Available from: https://ecdc.europa.eu/en/publications-data/hiv-modelling-tool.

95. van Sighem, A., et al., *Estimating HIV Incidence, Time to Diagnosis, and the Undiagnosed HIV Epidemic Using Routine Surveillance Data.* Epidemiology, 2015. **26**(5): p. 653-60.

96. Amundsen, E.J., et al., *Back-calculation based on HIV and AIDS registers in Denmark, Norway and Sweden 1977-95 among homosexual men: estimation of absolute rates, incidence rates and prevalence of HIV.* J Epidemiol Biostat, 2000. **5**(4): p. 233-43.

97. Hamers, F.F. and A.N. Phillips, *Diagnosed and undiagnosed HIV-infected populations in Europe.* HIV Med, 2008. **9 Suppl 2**: p. 6-12.

98. Pharris, A., et al., *Estimating HIV incidence and number of undiagnosed individuals living with HIV in the European Union/European Economic Area, 2015.* Euro Surveill, 2016. **21**(48).

99. Andersson, E., et al., *Challenges in modelling the proportion of undiagnosed HIV infections in Sweden.* Euro Surveill, 2019. **24**(14).

100. Giardina, F., et al., *Getting more from heterogeneous HIV-1 surveillance data in a high immigration country: estimation of incidence and undiagnosed population size using multiple biomarkers.* Int J Epidemiol, 2019.

101. Lodi, S., et al., *Time from human immunodeficiency virus seroconversion to reaching CD4+ cell count thresholds <200, <350, and <500 Cells/mm(3): assessment of need following changes in treatment guidelines.* Clin Infect Dis, 2011. **53**(8): p. 817-25.

102. Gaines, H., et al., *Immunological changes in primary HIV-1 infection.* Aids, 1990. **4**(10): p. 995-9.

103. Busch, M.P., et al., *Beyond detuning: 10 years of progress and new challenges in the development and application of assays for HIV incidence estimation.* Aids, 2010. **24**(18): p. 2763-71.

104. Mastro, T.D., et al., *Estimating HIV Incidence in Populations Using Tests for Recent Infection: Issues, Challenges and the Way Forward.* J HIV AIDS Surveill Epidemiol, 2010. **2**(1): p. 1-14.

105. Skar, H., J. Albert, and T. Leitner, *Towards estimation of HIV-1 date of infection: a time-continuous IgG-model shows that seroconversion does not occur at the midpoint between negative and positive tests.* PLoS One, 2013. **8**(4): p. e60906.

106. Parekh, B.S., et al., *Determination of mean recency period for estimation of HIV type 1 Incidence with the BED-capture EIA in persons infected with diverse subtypes.* AIDS Res Hum Retroviruses, 2011. **27**(3): p. 265-73.

107. Sexton, C.J., et al., *Correlation of prospective and cross-sectional measures of HIV type 1 incidence in a higher-risk cohort in Ho Chi Minh City, Vietnam.* AIDS Res Hum Retroviruses, 2012. **28**(8): p. 866-73.

108. Moyo, S., et al., *Identifying Recent HIV Infections: From Serological Assays to Genomics.* Viruses, 2015. **7**(10): p. 5508-24.

109. Duong, Y.T., et al., *Recalibration of the Limiting Antigen Avidity EIA to Determine Mean Duration of Recent Infection in Divergent HIV-1 Subtypes.* PLoS One, 2015. **10**(2).

110. Kassanjee, R., et al., *Viral load criteria and threshold optimization to improve HIV incidence assay characteristics.* Aids, 2016. **30**(15): p. 2361-71.

111. Andersson, E., et al., *Evaluation of sequence ambiguities of the HIV-1 pol gene as a method to identify recent HIV-1 infection in transmitted drug resistance surveys.* Infection Genetics and Evolution, 2013. **18**: p. 125-131.

112. Ragonnet-Cronin, M., et al., *Genetic diversity as a marker for timing infection in HIV-infected patients: evaluation of a 6-month window and comparison with BED.* J Infect Dis, 2012. **206**(5): p. 756-64.

113. Kouyos, R.D., et al., *Ambiguous nucleotide calls from population-based sequencing of HIV-1 are a marker for viral diversity and the age of infection.* Clin Infect Dis, 2011. **52**(4): p. 532-9.

114. Carlisle, L.A., et al., *Viral Diversity Based on Next-Generation Sequencing of HIV-1 Provides Precise Estimates of Infection Recency and Time Since Infection.* J Infect Dis, 2019. **220**(2): p. 254-265.

115. Park, S.Y., et al., *Developing high-throughput HIV incidence assay with pyrosequencing platform.* J Virol, 2014. **88**(5): p. 2977-90.

116. Puller, V., R. Neher, and J. Albert, *Estimating time of HIV-1 infection from next-generation sequence diversity.* PLoS Comput Biol, 2017. **13**(10): p. e1005775.

117. Park, S.Y., et al., *HIITE: HIV-1 incidence and infection time estimator.* Bioinformatics, 2018. **34**(12): p. 2046-2052.

118. Rice, B.D., et al., *A new method to assign country of HIV infection among heterosexuals born abroad and diagnosed with HIV.* Aids, 2012. **26**(15): p. 1961-6.

119. Cousins, M.M., et al., *A comparison of two measures of HIV diversity in multi-assay algorithms for HIV incidence estimation.* PLoS One, 2014. **9**(6): p. e101043.

120. Laeyendecker, O., et al., *Identification and validation of a multi-assay algorithm for cross-sectional HIV incidence estimation in populations with subtype C infection.* J Int AIDS Soc, 2018. **21**(2).

121. ECDC/WHO, *HIV/AIDS Surveillance in Europe 2018– 2017 data*. 2018: https://ecdc.europa.eu/sites/portal/files/documents/hiv-aids-surveillance-europe-2018.pdf.

122. Desgrees-du-Lou, A., et al., *Sub-Saharan African migrants living with HIV acquired after migration, France, ANRS PARCOURS study, 2012 to 2013.* Euro Surveill, 2015. **20**(46).

123. Bergdahl, S., et al., *Declining levels of HIV P24 antigen in serum during treatment with foscarnet.* Lancet, 1988. **1**(8593): p. 1052.

124. Bergdahl, S., et al., *Pronounced anti-HIV-1 activity of foscarnet in patients without cytomegalovirus infection.* J Acquir Immune Defic Syndr Hum Retrovirol, 1998. **18**(1): p. 51-3.

125. Julander, I., et al., *Treatment of HIV-1 infected patients with Peptide T* Antiviral Chemistry and Chemotherapy 1990. **1**(6): p. 349-354.

126. Berglund, O., et al., *Combined treatment of symptomatic human immunodeficiency virus type 1 infection with native interferon-alpha and zidovudine.* J Infect Dis, 1991. **163**(4): p. 710-5.

127. Parks, W.P., et al., *HIV-1 inhibition by azidothymidine in a concurrently randomized placebo-controlled trail.* J Acquir Immune Defic Syndr, 1988. **1**(2): p. 125-30.

128. Mitsuya, H., et al., *3'-Azido-3'-deoxythymidine (BW A509U): an antiviral agent that inhibits the infectivity and cytopathic effect of human T-lymphotropic virus type III/lymphadenopathy-associated virus in vitro.* Proc Natl Acad Sci U S A, 1985. **82**(20): p. 7096-100.

129. Fischl, M.A., et al., *The efficacy of azidothymidine (AZT) in the treatment of patients with AIDS and AIDS-related complex. A double-blind, placebo-controlled trial.* N Engl J Med, 1987. **317**(4): p. 185-91.

130. Larder, B.A., G. Darby, and D.D. Richman, *HIV with reduced sensitivity to zidovudine (AZT) isolated during prolonged therapy.* Science, 1989. **243**(4899): p. 1731-4.

131. Larder, B.A. and S.D. Kemp, *Multiple mutations in HIV-1 reverse transcriptase confer high-level resistance to zidovudine (AZT).* Science, 1989. **246**(4934): p. 1155-8.

132. Ridky, T. and J. Leis, *Development of drug resistance to HIV-1 protease inhibitors.* J Biol Chem, 1995. **270**(50): p. 29621-3.

133. Demeter, L.M., et al., *HIV-1 drug susceptibilities and reverse transcriptase mutations in patients receiving combination therapy with didanosine and delavirdine.* J Acquir Immune Defic Syndr Hum Retrovirol, 1997. **14**(2): p. 136-44.

134. Balzarini, J., et al., *Concomitant combination therapy for HIV infection preferable over sequential therapy with 3TC and non-nucleoside reverse transcriptase inhibitors.* Proc Natl Acad Sci U S A, 1996. **93**(23): p. 13152-7.

135. Schmit, J.C., et al., *Multiple drug resistance to nucleoside analogues and nonnucleoside reverse transcriptase inhibitors in an efficiently replicating human immunodeficiency virus type 1 patient strain.* J Infect Dis, 1996. **174**(5): p. 962-8.

136. Gulick, R.M., et al., *Treatment with indinavir, zidovudine, and lamivudine in adults with human immunodeficiency virus infection and prior antiretroviral therapy.* N Engl J Med, 1997. **337**(11): p. 734-9.

137. Hammer, S.M., 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, 1997. **337**(11): p. 725-33.

138. Collier, A.C., et al., *Treatment of human immunodeficiency virus infection with saquinavir, zidovudine, and zalcitabine. AIDS Clinical Trials Group.* N Engl J Med, 1996. **334**(16): p. 1011-7.

139. Mellors, J.W., et al., *Prognosis in HIV-1 infection predicted by the quantity of virus in plasma.* Science, 1996. **272**(5265): p. 1167-70.

140. Marschner, I.C., et al., *Use of changes in plasma levels of human immunodeficiency virus type 1 RNA to assess the clinical benefit of antiretroviral therapy.* J Infect Dis, 1998. **177**(1): p. 40-7.

141. Schmit, J.C. and B. Weber, *Recent advances in antiretroviral therapy and HIV infection monitoring.* Intervirology, 1997. **40**(5-6): p. 304-21.

142. Fischl, M.A., *Antiretroviral therapy in 1999 for antiretroviral-naive individuals with HIV infection.* Aids, 1999. **13 Suppl 1**: p. S49-59.

143. Gazzard, B. and G. Moyle, *1998 revision to the British HIV Association guidelines for antiretroviral treatment of HIV seropositive individuals. BHIVA Guidelines Writing Committee.* Lancet, 1998. **352**(9124): p. 314-6.

144. Carpenter, C.C., et al., *Antiretroviral therapy for HIV infection in 1998: updated recommendations of the International AIDS Society-USA Panel.* Jama, 1998. **280**(1): p. 78-86.

145. Moyle, G.J. and D. Back, *Principles and practice of HIV-protease inhibitor pharmacoenhancement.* HIV Med, 2001. **2**(2): p. 105-13.

146. Referensgruppen för antiviral terapi, *Antiretroviral behandling av hivinfektion − Behandlingsrekommendation 2019*. 2019: https://www.sls.se/rav/rekommendationer/hiv/Antiretroviral-behandling-av-hivinfektion-2019/. In Swedish.

147. Singh, K., et al., *GS-CA Compounds: First-In-Class HIV-1 Capsid Inhibitors Covering Multiple Grounds.* Front Microbiol, 2019. **10**: p. 1227.

148. Di Perri, G., et al., *The clinical pharmacology of integrase inhibitors.* Expert Rev Clin Pharmacol, 2018.

149. Mofenson, L.M., et al., *Optimizing responses to drug safety signals in pregnancy: the example of dolutegravir and neural tube defects.* J Int AIDS Soc, 2019. **22**(7): p. e25352.

150. Bourgi, K., et al., *Greater Weight Gain in Treatment Naive Persons Starting Dolutegravir-Based Antiretroviral Therapy.* Clin Infect Dis, 2019.

151. O'Brien, W.A., et al., *Changes in plasma HIV-1 RNA and CD4+ lymphocyte counts and the risk of progression to AIDS. Veterans Affairs Cooperative Study Group on AIDS.* N Engl J Med, 1996. **334**(7): p. 426-31.

152. U.S Department of Health and Human Services. *NIH aidsinfo*. 2018 [cited 2018 December, 10].

153. Bachmann, N., et al., *Importance of routine viral load monitoring: higher levels of resistance at ART failure in Uganda and Lesotho compared with Switzerland.* J Antimicrob Chemother, 2018.

154. Clutter, D.S., et al., *HIV-1 drug resistance and resistance testing.* Infect Genet Evol, 2016. **46**: p. 292-307.

155. Gupta, R.K., et al., *Virological monitoring and resistance to first-line highly active antiretroviral therapy in adults infected with HIV-1 treated under WHO guidelines: a systematic review and meta-analysis.* Lancet Infect Dis, 2009. **9**(7): p. 409-17.

156. Nettles, R.E., et al., *Intermittent HIV-1 viremia (Blips) and drug resistance in patients receiving HAART.* Jama, 2005. **293**(7): p. 817-29.

157. Sorstedt, E., et al., *Viral blips during suppressive antiretroviral treatment are associated with high baseline HIV-1 RNA levels.* BMC Infect Dis, 2016. **16**: p. 305.

158. Gonzalez-Serna, A., et al., *Performance of HIV-1 drug resistance testing at low-level viremia and its ability to predict future virologic outcomes and viral evolution in treatment-naive individuals.* Clin Infect Dis, 2014. **58**(8): p. 1165-73.

159. Gunthard, H.F., et al., *Human Immunodeficiency Virus Drug Resistance: 2018 Recommendations of the International Antiviral Society-USA Panel.* Clin Infect Dis, 2019. **68**(2): p. 177-187.

160. Boucher, C.A., et al., *State of the Art in HIV Drug Resistance: Science and Technology Knowledge Gap.* AIDS Rev, 2018. **20**(1): p. 27-42.

161. Bachmann, N., et al., *Determinants of HIV-1 reservoir size and long-term dynamics during suppressive ART.* Nat Commun, 2019. **10**(1): p. 3193.

162. Venner, C.M., et al., *Infecting HIV-1 Subtype Predicts Disease Progression in Women of Sub-Saharan Africa.* EBioMedicine, 2016. **13**: p. 305-314.

163. Amornkul, P.N., et al., *Disease progression by infecting HIV-1 subtype in a seroconverter cohort in sub-Saharan Africa.* Aids, 2013. **27**(17): p. 2775-86.

164. *Mortality of treated HIV-1 positive individuals according to viral subtype in Europe and Canada: collaborative cohort analysis.* Aids, 2016. **30**(3): p. 503-13.

165. Ssemwanga, D., et al., *Effect of HIV-1 subtypes on disease progression in rural Uganda: a prospective clinical cohort study.* PLoS One, 2013. **8**(8): p. e71768.

166. Pant Pai, N., S. Shivkumar, and J.M. Cajas, *Does genetic diversity of HIV-1 non-B subtypes differentially impact disease progression in treatment-naive HIV-1-infected individuals? A systematic review of evidence: 1996-2010.* J Acquir Immune Defic Syndr, 2012. **59**(4): p. 382-8.

167. Tatarelli, P., et al., *HIV-1 RNA quantification in CRF02_AG HIV-1 infection: too easy to make mistakes.* New Microbiol, 2016. **39**(2): p. 150-2.

168. Gueudin, M., et al., *Evaluation of the Roche Cobas TaqMan and Abbott RealTime extraction-quantification systems for HIV-1 subtypes.* J Acquir Immune Defic Syndr, 2007. **44**(5): p. 500-5.

169. Church, D., et al., *Comparison of the RealTime HIV-1, COBAS TaqMan 48 v1.0, Easy Q v1.2, and Versant v3.0 assays for determination of HIV-1 viral loads in a cohort of Canadian patients with diverse HIV subtype infections.* J Clin Microbiol, 2011. **49**(1): p. 118-24.

170. Alaeus, A., et al., *Subtype-specific problems with quantification of plasma HIV-1 RNA.* Aids, 1997. **11**(7): p. 859-65.

171. Bhargava, M., et al., *Do HIV-1 non-B subtypes differentially impact resistance mutations and clinical disease progression in treated populations? Evidence from a systematic review.* J Int AIDS Soc, 2014. **17**: p. 18944.

172. Atlas, A., et al., *Impact of HIV type 1 genetic subtype on the outcome of antiretroviral therapy.* AIDS Res Hum Retroviruses, 2005. **21**(3): p. 221-7.

173. Haggblom, A., et al., *Virological failure in patients with HIV-1 subtype C receiving antiretroviral therapy: an analysis of a prospective national cohort in Sweden.* Lancet HIV, 2016. **3**(4): p. e166-74.

174. Kantor, R., et al., *Pretreatment HIV Drug Resistance and HIV-1 Subtype C Are Independently Associated With Virologic Failure: Results From the Multinational PEARLS (ACTG A5175) Clinical Trial.* Clin Infect Dis, 2015. **60**(10): p. 1541-9.

175. Neogi, U., et al., *Novel tetra-peptide insertion in Gag-p6 ALIX-binding motif in HIV-1 subtype C associated with protease inhibitor failure in Indian patients.* Aids, 2014. **28**(15): p. 2319-22.

176. Neogi, U., et al., *Mutational Heterogeneity in p6 Gag Late Assembly (L) Domains in HIV-1 Subtype C Viruses from South Africa.* AIDS Res Hum Retroviruses, 2016. **32**(1): p. 80-4.

177. Aralaguppe, S.G., et al., *Increased replication capacity following evolution of PYxE insertion in Gag-p6 is associated with enhanced virulence in HIV-1 subtype C from East Africa.* J Med Virol, 2017. **89**(1): p. 106-111.

178. van Domselaar, R., et al., *HIV-1 Subtype C with PYxE Insertion Has Enhanced Binding of Gag-p6 to Host Cell Protein ALIX and Increased Replication Fitness.* J Virol, 2019. **93**(9).

179. Sutherland, K.A., et al., *HIV-1 subtype influences susceptibility and response to monotherapy with the protease inhibitor lopinavir/ritonavir.* J Antimicrob Chemother, 2015. **70**(1): p. 243-8.

180. Njenda, D.T., et al., *Antiretroviral potency of 4'-ethnyl-2'-fluoro-2'-deoxyadenosine, tenofovir alafenamide and second-generation NNRTIs across diverse HIV-1 subtypes.* J Antimicrob Chemother, 2018. **73**(10): p. 2721-2728.

181. Neogi, U., et al., *Factors influencing the efficacy of rilpivirine in HIV-1 subtype C in low- and middle-income countries.* J Antimicrob Chemother, 2016. **71**(2): p. 367-71.

182. Hill, K.J., et al., *Strain-specific effect on biphasic DNA binding by HIV-1 integrase.* Aids, 2018.

183. Telele, N.F., et al., *Pretreatment drug resistance in a large countrywide Ethiopian HIV-1C cohort: a comparison of Sanger and high-throughput sequencing.* Sci Rep, 2018. **8**(1): p. 7556.

184. Rogers, L., et al., *Structural Implications of Genotypic Variations in HIV-1 Integrase From Diverse Subtypes.* Front Microbiol, 2018. **9**: p. 1754.

185. Overton, E.O., C.; Swindells, S.; Arasteh, K.; et al, *Monthly Long-Acting Cabotegravir and Rilpivirine is Noninferior to Oral ART as Maintenance Therapy for HIV-1 Infection: Week 48 Pooled Analysis From the Phase 3 ATLAS and FLAIR Studies*, in *10th IAS Conference on HIV Science; July 21–24, 2019*. 2019: Mexico City,Mexico. Poster.

186. Soriano, V., et al., *Dual antiretroviral therapy for HIV infection.* Expert Opin Drug Saf, 2017. **16**(8): p. 923-932.

187. Cortez, K.J. and F. Maldarelli, *Clinical management of HIV drug resistance.* Viruses, 2011. **3**(4): p. 347-78.

188. Condra, J.H., *Resistance to HIV protease inhibitors.* Haemophilia, 1998. **4**(4): p. 610-5.

189. Wensing, A.M., et al., *2017 Update of the Drug Resistance Mutations in HIV-1.* Top Antivir Med, 2017. **24**(4): p. 132-133.

190. Nijhuis, M., et al., *Increased fitness of drug resistant HIV-1 protease as a result of acquisition of compensatory mutations during suboptimal therapy.* Aids, 1999. **13**(17): p. 2349-59.

191. Pedersen, O.S. and E.B. Pedersen, *Non-nucleoside reverse transcriptase inhibitors: the NNRTI boom.* Antivir Chem Chemother, 1999. **10**(6): p. 285-314.

192. Gardner, E.M., et al., *Antiretroviral medication adherence and class- specific resistance in a large prospective clinical trial.* Aids, 2010. **24**(3): p. 395-403.

193. Cozzi-Lepri, A., et al., *Thymidine analogue mutation profiles: factors associated with acquiring specific profiles and their impact on the virological response to therapy.* Antivir Ther, 2005. **10**(7): p. 791-802.

194. Kityo, C., et al., *HIV Drug Resistance Mutations in Non-B Subtypes After Prolonged Virological Failure on NNRTI-Based First-Line Regimens in Sub-Saharan Africa.* J Acquir Immune Defic Syndr, 2017. **75**(2): p. e45-e54.

195. Koning, F.A., et al., *Subtype-specific differences in the development of accessory mutations associated with high-level resistance to HIV-1 nucleoside reverse transcriptase inhibitors.* J Antimicrob Chemother, 2013. **68**(6): p. 1220-36.

196. Blanco, J.L., et al., *HIV-1 integrase inhibitor resistance and its clinical implications.* J Infect Dis, 2011. **203**(9): p. 1204-14.

197. Neogi, U., et al., *Ex-vivo antiretroviral potency of newer integrase strand transfer inhibitors cabotegravir and bictegravir in HIV type 1 non-B subtypes.* Aids, 2018. **32**(4): p. 469-476.

198. Wijting, I., et al., *Dolutegravir as maintenance monotherapy for HIV (DOMONO): a phase 2, randomised non-inferiority trial.* Lancet HIV, 2017. **4**(12): p. e547-e554.

199. Gunthard, H.F., et al., *Human Immunodeficiency Virus Drug Resistance: 2018 Recommendations of the International Antiviral Society-USA Panel.* Clin Infect Dis, 2018.

200. Tang, M.W. and R.W. Shafer, *HIV-1 antiretroviral resistance: scientific principles and clinical applications.* Drugs, 2012. **72**(9): p. e1-25.

201. Das, K. and E. Arnold, *HIV-1 reverse transcriptase and antiviral drug resistance. Part 2.* Curr Opin Virol, 2013. **3**(2): p. 119-28.

202. Reece, R., et al., *Accumulated pre-switch resistance to more recently introduced one-pill-once-a-day antiretroviral regimens impacts HIV-1 virologic outcome.* J Clin Virol, 2018. **105**: p. 11-17.

203. Erice, A., et al., *Brief report: primary infection with zidovudine-resistant human immunodeficiency virus type 1.* N Engl J Med, 1993. **328**(16): p. 1163-5.

204. Birk, M. and A. Sonnerborg, *Variations in HIV-1 pol gene associated with reduced sensitivity to antiretroviral drugs in treatment-naive patients.* Aids, 1998. **12**(18): p. 2369-75.

205. Wahlberg, J., et al., *Apparent selection against transmission of zidovudine-resistant human immunodeficiency virus type 1 variants.* J Infect Dis, 1994. **169**(3): p. 611-4.

206. Sonnerburg, A., et al., *Transmission of zidovudine-resistant HIV-1.* Aids, 1993. **7**(12): p. 1684-5.

207. Wittkop, L., et al., *Effect of transmitted drug resistance on virological and immunological response to initial combination antiretroviral therapy for HIV (EuroCoord-CHAIN joint project): a European multicohort study.* Lancet Infect Dis, 2011. **11**(5): p. 363-71.

208. Hamers, R.L., et al., *Effect of pretreatment HIV-1 drug resistance on immunological, virological, and drug-resistance outcomes of first-line antiretroviral treatment in sub-Saharan Africa: a multicentre cohort study.* Lancet Infect Dis, 2012. **12**(4): p. 307-17.

209. Castro, H., et al., *Persistence of HIV-1 transmitted drug resistance mutations.* J Infect Dis, 2013. **208**(9): p. 1459-63.

210. Wertheim, J.O., et al., *Transmission fitness of drug-resistant HIV revealed in a surveillance system transmission network.* Virus Evol, 2017. **3**(1): p. vex008.

211. Winand, R., et al., *Assessing transmissibility of HIV-1 drug resistance mutations from treated and from drug-naive individuals.* Aids, 2015. **29**(15): p. 2045-52.

212. Kuhnert, D., et al., *Quantifying the fitness cost of HIV-1 drug resistance mutations through phylodynamics.* PLoS Pathog, 2018. **14**(2): p. e1006895.

213. Mbisa, J.L., et al., *Evidence of Self-Sustaining Drug Resistant HIV-1 Lineages Among Untreated Patients in the United Kingdom.* Clin Infect Dis, 2015. **61**(5): p. 829-36.

214. Vega, Y., et al., *Epidemiological Surveillance of HIV-1 Transmitted Drug Resistance in Spain in 2004-2012: Relevance of Transmission Clusters in the Propagation of Resistance Mutations.* PLoS One, 2015. **10**(5): p. e0125699.

215. Paraskevis, D., et al., *Prevalence of drug resistance among HIV-1 treatment-naive patients in Greece during 2003-2015: Transmitted drug resistance is due to onward transmissions.* Infect Genet Evol, 2017. **54**: p. 183-191.

216. Drescher, S.M., et al., *Treatment-naive individuals are the major source of transmitted HIV-1 drug resistance in men who have sex with men in the Swiss HIV Cohort Study.* Clin Infect Dis, 2014. **58**(2): p. 285-94.

217. Boucher, C.A., et al., *State of the Art in HIV Drug Resistance: Surveillance and Regional Gaps.* AIDS Rev, 2018. **20**(1): p. 43-57.

218. Rhee, S.Y., et al., *Geographic and temporal trends in the molecular epidemiology and genetic mechanisms of transmitted HIV-1 drug resistance: an individual-patient- and sequence-level meta-analysis.* PLoS Med, 2015. **12**(4): p. e1001810.

219. Arimide, D.A., et al., *HIV-genetic diversity and drug resistance transmission clusters in Gondar, Northern Ethiopia, 2003-2013.* PLoS One, 2018. **13**(10): p. e0205446.

220. McGee, K.S., et al., *Canary in the Coal Mine? Transmitted Mutations Conferring Resistance to All Integrase Strand Transfer Inhibitors in a Treatment-Naive Patient.* Open Forum Infect Dis, 2018. **5**(11): p. ofy294.

221. Boyd, S.D., et al., *Transmitted raltegravir resistance in an HIV-1 CRF_AG-infected patient.* Antivir Ther, 2011. **16**(2): p. 257-61.

222. Young, B., et al., *Transmission of integrase strand-transfer inhibitor multidrug-resistant HIV-1: case report and response to raltegravir-containing antiretroviral therapy.* Antivir Ther, 2011. **16**(2): p. 253-6.

223. World Health Organization, *Surveillance of HIV drug resistance in adults initiating antiretroviral therapy (pretreatment HIV drug resistance)*. 2014: https://www.who.int/hiv/topics/drugresistance/protocols/en/.

224. Heger, E., et al., *Development of a phenotypic susceptibility assay for HIV-1 integrase inhibitors.* J Virol Methods, 2016. **238**: p. 29-37.

225. Palmer, S., et al., *Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis.* J Clin Microbiol, 2005. **43**(1): p. 406-13.

226. Larder, B.A., et al., *Quantitative detection of HIV-1 drug resistance mutations by automated DNA sequencing.* Nature, 1993. **365**(6447): p. 671-3.

227. Church, J.D., et al., *Sensitivity of the ViroSeq HIV-1 genotyping system for detection of the K103N resistance mutation in HIV-1 subtypes A, C, and D.* J Mol Diagn, 2006. **8**(4): p. 430-2; quiz 527.

228. Leitner, T., et al., *Analysis of heterogeneous viral populations by direct DNA sequencing.* Biotechniques, 1993. **15**(1): p. 120-7.

229. Inzaule, S.C., et al., *Clinically relevant thresholds for ultrasensitive HIV drug resistance testing: a multi-country nested case-control study*. Lancet HIV, 2018. **5**(11): p. e638-e646.

230. Cozzi-Lepri, A., et al., *Low-frequency drug-resistant HIV-1 and risk of virological failure to first-line NNRTI-based ART: a multicohort European case-control study using centralized ultrasensitive 454 pyrosequencing.* J Antimicrob Chemother, 2015. **70**(3): p. 930-40.

231. Li, J.Z., et al., *Low-frequency HIV-1 drug resistance mutations and risk of NNRTI-based antiretroviral treatment failure: a systematic review and pooled analysis.* Jama, 2011. **305**(13): p. 1327-35.

232. Ji, H., et al., *Bioinformatic data processing pipelines in support of next-generation sequencing-based HIV drug resistance testing: the Winnipeg Consensus.* J Int AIDS Soc, 2018. **21**(10): p. e25193.

233. Derache, A., et al., *Impact of Next-generation Sequencing Defined Human Immunodeficiency Virus Pretreatment Drug Resistance on Virological Outcomes in the ANRS 12249 Treatment-as-Prevention Trial.* Clin Infect Dis, 2019. **69**(2): p. 207-214.

234. Metzner, K.J., et al., *Efficient suppression of minority drug-resistant HIV type 1 (HIV-1) variants present at primary HIV-1 infection by ritonavir-boosted protease inhibitor-containing antiretroviral therapy.* J Infect Dis, 2010. **201**(7): p. 1063-71.

235. Nguyen, T., et al., *Prevalence and clinical impact of minority resistant variants in patients failing an integrase inhibitor-based regimen by ultra-deep sequencing.* J Antimicrob Chemother, 2018. **73**(9): p. 2485-2492.

236. Parry, C.M., et al., *Field study of dried blood spot specimens for HIV-1 drug resistance genotyping.* J Clin Microbiol, 2014. **52**(8): p. 2868-75.

237. Svard, J., et al., *Drug resistance testing through remote genotyping and predicted treatment options in human immunodeficiency virus type 1 infected Tanzanian subjects failing first or second line antiretroviral therapy.* PLoS One, 2017. **12**(6): p. e0178942.

238. Liu, T.F. and R.W. Shafer, *Web resources for HIV type 1 genotypic-resistance test interpretation.* Clin Infect Dis, 2006. **42**(11): p. 1608-18.

239. Zazzi, M., et al., *Predicting response to antiretroviral treatment by machine learning: the EuResist project.* Intervirology, 2012. **55**(2): p. 123-7.

240. Lawyer, G., et al., *HIV-1 mutational pathways under multidrug therapy.* AIDS Res Ther, 2011. **8**: p. 26.

241. Altmann, A., et al., *Predicting the response to combination antiretroviral therapy: retrospective validation of geno2pheno-THEO on a large clinical database.* J Infect Dis, 2009. **199**(7): p. 999-1006.

242. Altmann, A., et al., *Advantages of predicted phenotypes and statistical learning models in inferring virological response to antiretroviral therapy from HIV genotype.* Antivir Ther, 2009. **14**(2): p. 273-83.

243. Altmann, A., et al., *Comparison of classifier fusion methods for predicting response to anti HIV-1 therapy.* PLoS One, 2008. **3**(10): p. e3470.

244. Rosen-Zvi, M., et al., *Selecting anti-HIV therapies based on a variety of genomic and clinical factors.* Bioinformatics, 2008. **24**(13): p. i399-406.

245. Zazzi, M., et al., *Prediction of response to antiretroviral therapy by human experts and by the EuResist data-driven expert system (the EVE study).* HIV Med, 2011. **12**(4): p. 211-8.

246. Rhee, S.Y., et al., *HIV-1 Drug Resistance Mutations: Potential Applications for Point-of-Care Genotypic Resistance Testing.* PLoS One, 2015. **10**(12): p. e0145772.

247. Clutter, D.S., et al., *Genetic Variability of HIV-1 for Drug Resistance Assay Development.* Viruses, 2016. **8**(2).

248. Duarte, H.A., et al., *Current Status of Point-of-Care Testing for Human Immunodeficiency Virus Drug Resistance.* J Infect Dis, 2017. **216**(suppl_9): p. S824-s828.

249. Gupta, R.K., et al., *HIV-1 drug resistance before initiation or re-initiation of first-line antiretroviral therapy in low-income and middle-income countries: a systematic review and meta-regression analysis.* Lancet Infect Dis, 2018. **18**(3): p. 346-355.

250. Hosseinipour, M.C., et al., *Emergence of HIV drug resistance during first- and second-line antiretroviral therapy in resource-limited settings.* J Infect Dis, 2013. **207 Suppl 2**: p. S49-56.

251. Barth, R.E., et al., *Virological follow-up of adult patients in antiretroviral treatment programmes in sub-Saharan Africa: a systematic review.* Lancet Infect Dis, 2010. **10**(3): p. 155-66.

252. Boender, T.S., et al., *Pretreatment HIV drug resistance increases regimen switches in sub-Saharan Africa.* Clin Infect Dis, 2015. **61**(11): p. 1749-58.

253. World Health Organization, *Global action plan on HIV drug resistance 2017–2021. Geneva: World Health Organization; 2017. Licence: CC BY-NC-SA 3.0 IGO.* 2017: https://apps.who.int/iris/bitstream/handle/10665/255883/9789241512848-eng.pdf;jsessionid=61D7416EC19FBE751478E0DC1774104D?sequence=1.

254. Amstutz, A., et al., *SESOTHO trial ("Switch Either near Suppression Or THOusand") - switch to second-line versus WHO-guided standard of care for unsuppressed patients on first-line ART with viremia below 1000 copies/mL: protocol of a multicenter, parallel-group, open-label, randomized clinical trial in Lesotho, Southern Africa.* BMC Infect Dis, 2018. **18**(1): p. 76.

255. Labhardt, N.D., et al., *Should viral load thresholds be lowered?: Revisiting the WHO definition for virologic failure in patients on antiretroviral therapy in resource-limited settings.* Medicine (Baltimore), 2016. **95**(28): p. e3985.

256. World Health Organization, Switzerland, *Update of recommendations on first- and second-line antiretroviral regimens.* 2019: www.who.int/hiv.

257. Bennett, D.E., et al., *Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 update.* PLoS One, 2009. **4**(3): p. e4724.

258. Gifford, R.J., et al., *The calibrated population resistance tool: standardized genotypic estimation of transmitted HIV-1 drug resistance.* Bioinformatics, 2009. **25**(9): p. 1197-8.

259. van de Laar, M.J., et al., *Piloting a surveillance system for HIV drug resistance in the European Union.* Euro Surveill, 2019. **24**(19).

260. Hofstra, L.M., et al., *Transmission of HIV Drug Resistance and the Predicted Effect on Current First-line Regimens in Europe.* Clin Infect Dis, 2016. **62**(5): p. 655-63.

261. Keating, P., et al., *Assessment of HIV molecular surveillance capacity in the European Union, 2016.* Euro Surveill, 2017. **22**(49).

262. Gupta, R.K., et al., *Global trends in antiretroviral resistance in treatment-naive individuals with HIV after rollout of antiretroviral treatment in resource-limited settings: a global collaborative study and meta-regression analysis.* Lancet, 2012. **380**(9849): p. 1250-8.

263. World Health Organization, *World Health Organization. HIV Drug Resistance Report 2017.* 2017: https://www.who.int/hiv/pub/drugresistance/hivdr-report-2017/en/.

264. Karlsson, A., et al., *Low prevalence of transmitted drug resistance in patients newly diagnosed with HIV-1 infection in Sweden 2003-2010.* PLoS One, 2012. **7**(3): p. e33484.

265. Andersson, E., et al., *Increase in transmitted drug resistance in migrants from sub-Saharan Africa diagnosed with HIV-1 in Sweden.* AIDS, 2018. **32**(7): p. 877-884.

266. Phillips, A.N., et al., *Impact of HIV Drug Resistance on HIV/AIDS-Associated Mortality, New Infections, and Antiretroviral Therapy Program Costs in Sub-Saharan Africa.* J Infect Dis, 2017. **215**(9): p. 1362-1365.

267. Alcantara, L.C., et al., *A standardized framework for accurate, high-throughput genotyping of recombinant and non-recombinant viral sequences.* Nucleic Acids Res, 2009. **37**(Web Server issue): p. W634-42.

268. de Oliveira, T., et al., *An automated genotyping system for analysis of HIV-1 and other microbial sequences.* Bioinformatics, 2005. **21**(19): p. 3797-800.

269. *Los Alamos Sequence Database*. www.hiv.lanl.gov.

270. Guindon, S., et al., *New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0.* Syst Biol, 2010. **59**(3): p. 307-21.

271. Hordijk, W. and O. Gascuel, *Improving the efficiency of SPR moves in phylogenetic tree search methods based on maximum likelihood.* Bioinformatics, 2005. **21**(24): p. 4338-47.

272. Anisimova, M. and O. Gascuel, *Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative.* Syst Biol, 2006. **55**(4): p. 539-52.

273. Ragonnet-Cronin, M., et al., *Automated analysis of phylogenetic clusters.* BMC Bioinformatics, 2013. **14**: p. 317.

274. Hassan, A.S., et al., *Defining HIV-1 transmission clusters based on sequence data.* AIDS, 2017. **31**(9): p. 1211-1222.

275. Aralaguppe, S.G., et al., *Multiplexed next-generation sequencing and de novo assembly to obtain near full-length HIV-1 genome from plasma virus.* J Virol Methods, 2016. **236**: p. 98-104.

276. Aralaguppe, S.G., et al., *MiDRMpol: A High-Throughput Multiplexed Amplicon Sequencing Workflow to Quantify HIV-1 Drug Resistance Mutations against Protease, Reverse Transcriptase, and Integrase Inhibitors.* Viruses, 2019. **11**(9).