The association between HIV infection duration/biomarkers of HIV infection and serious non-AIDS events (SNAEs)

Alexandra Rachel Campbell Lyons

Thesis manuscript submitted for the degree of Doctor of Philosophy

University College London

Declaration

I, Alexandra Rachel Campbell Lyons confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Life-long combination antiretroviral therapy (cART) has markedly increased life-expectancy in people living with HIV (PLHIV). Once largely AIDS-related, morbidity and mortality is now dominated by non-AIDS conditions including: cardiovascular, endocrine, renal and liver disease and non-AIDS defining cancers. These conditions, called Serious non-AIDS events (SNAEs), occur at higher rates in PLHIV than in the general population. This could be due to the effects of HIV infection, cART, other risk factors, or some combination of these. My study aimed to disentangle, as far as is possible using regression modelling of observational data, the relative contribution of each component to several SNAEs.

I explored whether HIV-related exposures were independently associated with fractures, diabetes mellitus (DM) and myocardial infarction (MI). I analysed data from the Concerted Action on Seroconversion to AIDS and Death in Europe (CASCADE) cohort collaboration. All individuals had well-estimated HIV-seroconversion dates. Exposures of interest comprised: duration of HIV infection, current and nadir CD4 cell count, current HIV viral load and duration of immune-suppression (≤200/100/50 cells/μL).

During 128 654 person-years of follow-up, 266 fractures, 103 MI and 109 DM (all first known events) occurred. I found evidence of an independent association between duration of HIV infection and MI, and between prior AIDS and fractures/MI, but evidence for other associations was lacking. After adjustment, each additional 10 years of HIV infection was associated with an approximate doubling in MI rate.

Initiation of cardiovascular disease risk reduction strategies at lower traditional risk thresholds may be warranted in PLHIV.

This is the first study to examine these associations in a large sample of individuals with well-estimated HIV-seroconversion dates, enabling accurate estimation of nadir CD4 and duration of both HIV infection and immune-suppression.

Acknowledgments

I would like to thank my supervisors Professor Caroline Sabin and Professor Kholoud Porter for their numerous explanations, appraising my work and for sharing their extensive knowledge of HIV epidemiology with me.

I am greatly indebted to Dr Deborah Ford for all her advice (especially regarding statistics) and her unfailing encouragement; without her input I would not have finished this thesis. I am also immensely grateful to Professor Ali Judd for her generosity in giving me leave to complete my thesis and for her very helpful comments and suggestions.

This work utilised data from the CASCADE cohort collaboration and many members of the CASCADE team and collaborating cohorts contributed to its completion. Many thanks to Ashely Olson for the time she spent answering my questions regarding CASCADE.

I am also grateful to the Principal Investigators and data managers of collaborating cohorts who participated in this project. Their willingness to share data and expertise has not only benefitted my small study, but numerous other collaborative projects across Europe and Worldwide which have substantially benefitted human health.

Impact Statement

PLHIV who receive treatment now have a life-expectancy approaching that of the general population. Despite this, previous studies suggest that they are at increased risks of serious non-AIDS events including: cardiovascular disease, non-AIDS cancers, DM, liver disease, kidney disease and fractures. These conditions are leading causes of ill-health and early death worldwide and treatment is often complex and expensive.

My study found that each additional decade of HIV infection is associated with a near doubling of MI incidence after adjusting for current age and other factors. This association was seen even in those with high current CD4 cell counts and undetectable viral loads. Ischaemic heart disease and HIV are predicted to be two of the three leading causes of morbidity worldwide by the end of the decade. Cardiovascular diseases are the most common cause of death in Western Europe. The HIV positive population is ageing. MI incidence increases with age. Therefore my findings are potentially important for future health service planning and policy. Whilst the possibility of residual confounding remains, if causal, my findings also have implications for the clinical management of the ageing HIV-positive population. Preventing CVD morbidity and mortality is dependent on accurate risk assessment.

My findings, suggest that CVD risk is being under-estimated in those with long-standing HIV. Implementation of risk-reduction measures at lower thresholds of traditional risk factors may be beneficial in people who have been living with HIV for a prolonged period, even if they have excellent HIV viraemic control.

I found little evidence that current HIV viral load or current/cumulative immune-compromise were independently associated with the incidence of fractures, DM or MI. I also found no evidence of an independent association between duration of HIV infection and fractures or DM. My study found that a prior AIDS diagnosis was associated with both fracture and MI incidence and that HCV-seropositivity was associated with both fractures and DM. These findings add to a body of existing work suggesting that both AIDS and HCV sero-positivity have multiple adverse effects.

My study provides valuable information regarding how the incidence of specific non-AIDS events is associated with HIV-related factors. Unlike most previous studies, all individuals in my analysis had a well-estimated date of HIV seroconversion. This enabled me to accurately estimate duration of both HIV infection and immune-suppression as well as nadir CD4 cell count, a major study strength. My study was observational, so causality

cannot be assumed. Data on a number of potential confounders including smoking were not available, which could have influenced my results.

I have presented my findings at international meetings and conferences, including an oral presentation at the European AIDS Clinical Society Conference, which was subsequently reported in the press. I plan to publish my findings in peer-reviewed journals.

Table of Contents

List of Tables

List of Figures

Abbreviations

AHIVCOS	Austrian HIV Cohort Study
AIDS	Acquired immunodeficiency syndrome
AMACS	Athens Multicenter AIDS Cohort Study
(c)ART	(combination) antiretroviral therapy
ART-CC	Antiretroviral Therapy Cohort Collaboration
ATHENA	AIDS Therapy Evaluation in the Netherlands
AQU	Aquitaine cohort
CASCADE	Concerted Action on SeroConversion to AIDS and Death in Europe
CHAMPS	Chronic Hepatitis C Management to Improve Outcomes study
CMV	Cytomegalovirus
CoRIS	Cohort of the Spanish HIV Research Network
D:A:D	The Data Collection on Adverse events of Anti-HIV Drugs
FHDH	The French Hospital Database on HIV (ANRS CO4)
FIRST	Flexible Initial Retrovirus Suppressive Therapies
HBV	Hepatitis B virus
HCV	Hepatitis C virus
HICDEP	HIV Cohorts Data Exchange Protocol
HIV	Human Immunodeficiency virus
HR	Hazard Ratio
HOPS	HIV Outpatient Study
ICONA	Italian Cohort of Antiretroviral Naïve Patients
INSIGHT	International Network for Strategic Initiatives in Global HIV Trials
IRR	Incident Rate Ratio
KP	Kaiser Permanente Study
MACS	The Multicentre AIDS Cohort Study
MASTER	Standardized Management of Antiviral Therapy
MFP	Multivariable fractional polynomials
MSM	Men who have sex with Men
NA-ACCORD	The North American AIDS Cohort Collaboration on Research and Design
NOR	Oslo and Ulleval hospital cohorts
NNRTI	Non-nucleotide reverse transcriptase inhibitors
NRTI	Nucleotide reverse transcriptase inhibitors
OR	Odds Ratio
PI	Protease Inhibitors
PLHIV	People living with HIV

PRIMO	Primary Infection Cohort (ANRS CO6)
PYFU	Person-years of follow-up
SAL	Southern Alberta Clinic Cohort
SEROCO	Seroconverter Cohort (ANRS CO2)
SHCS	The Swiss Hospital Cohort Study
SNAEs	Serious non-AIDS events
SUN	Study to Understand the Natural History of HIV/AIDS in the Era of Effective Therapy
UKR	The UK Register of HIV seroconverters
VACS-VC	The Veterans Aging Cohort-Virtual Cohort
WIHS	Women's Interagency HIV Study

Chapter 1: Introduction

1.1 Background

Human Immunodeficiency Virus (HIV) has had a substantial impact on the health of our species [1]. Untreated HIV infection leads, in nearly all cases, to progressive destruction of the immune system and death [2]. This destruction results in individuals becoming increasingly susceptible to certain tumours and opportunistic infections. The development of one of these conditions results in a diagnosis of Acquired Immunodeficiency Syndrome (AIDS) [3]. A list of opportunist infections and cancers which comprise the AIDS-defining illnesses can be found in Appendix A.

At the beginning of 2018 it was estimated that approximately 37 million people were currently living with HIV and a similar number had ever died from AIDS-related causes [4]. In those receiving no treatment, AIDS usually occurs about a decade after initial infection [5]. From then on, individuals generally have a poor prognosis with a median survival of less than two years if untreated [6]. There is currently no effective cure or vaccine, but since the introduction of combination antiretroviral therapy (cART) in 1996, there has been increasingly effective treatment [7]. A huge increase in life expectancy has been observed, in those with access to these drugs, over the last two decades [7]. Some people living with HIV (PLHIV) now have near normal life expectancy [7, 8]. These individuals typically have: been infected in more recent calendar years, started cART early in the course of infection, never inject drugs, been both treatment adherent and maintained or restore normal immune function [7-11].

It is only in the last twenty years, that individuals living with HIV have survived long enough to experience diseases associated with ageing. These include: cardiovascular disease, liver disease, kidney disease, fragility fractures, type-2 diabetes mellitus and non-AIDS defining cancers [12-17]. These events are often collectively referred to as Serious Non-AIDS Events or SNAEs [18-20] and are a major cause of morbidity and mortality [21-25]. There is evidence that many SNAEs occur more frequently in those with HIV than in the general population or in those without HIV infection (which would exclude those with HIV) [26-30].

1.2 HIV

1.2.1 Origins and discovery

The HIV pandemic originated in central Africa when the Simian Immunodeficiency Virus (SIV) crossed the species barrier from chimpanzees to humans [1]. This is estimated to have occurred sometime in the first 30 years of the 20th Century [1]. The virus came to the attention of the medical profession in the US at the beginning of the 1980s. Clinicians noted that cases of a number of previously rare conditions including Kaposi 's sarcoma and *Pneumocystis jiroveci* pneumonia (PJP) (at the time known as *Pneumocystis carinii* pneumonia (PCP)) were becoming more common in men who had sex with men (MSM) [31, 32]. These conditions were associated with an unusually high death rate which led to speculation that they represented an epidemic with a single cause [33]. In 1983 researchers correctly identified a retrovirus as the likely agent and this subsequently became known as HIV [34-36].

1.2.2 Classification

Two separate HIV viruses have so far been identified, HIV-1 and HIV-2, which are closely related lentiviruses [1]. HIV-1 is more common and virulent [1]. HIV-2 is mostly found in West Africa [1]. Both viruses are divided into a number of groups. The groups for HIV-1 are M (main), N, O and P. These groups are thought to correspond to separate incidents whereby SIV crossed the species barrier. Group M is very common whereas the others are rare [1]. Group M is responsible for the HIV pandemic and is further divided into subtypes A-K. There are also circulating recombinant forms (CRFs) where genetic material from different subtypes have combined [37]. B is the most common subtype in Western Europe, America and Australasia. It is the subtype most extensively studied [38]. Subtype C predominates in Southern Africa, Eastern Africa and India and accounts for nearly half of infections worldwide [37].

1.2.3 Transmission

HIV is transmitted through contact with certain bodily fluids. These comprise: semen, vaginal fluid, blood, breast-milk and serum from damaged mucosal membranes or skin. It is rare for someone to contract the virus through exposure to saliva, urine or tears [39]. Infection commonly occurs during: anal or vaginal intercourse, child birth, breastfeeding, the sharing of hypodermic needles, exposure to blood products or contaminated surgical instruments [39, 40].

1.2.4 Life-cycle

The lifecycle of the HIV virus involves a number of stages which are outlined below [2, 41-43]:

Viral Penetration (Binding and fusion): Once inside the body, HIV attacks cells expressing CD4 receptors. CD4 receptors are found at high concentrations on T-helper cells but are also present on monocytes, macrophages and dendritic cells [43]. The virus infects T-helper cells by attaching to the CD4 receptor and then binds to one of two co-receptors. A tiny proportion of individuals are partially or completely resistant to HIV, mostly due to a gene mutation [44]. After attaching, the virus envelope fuses with the cell membrane and releases its RNA and enzymes (including reverse transcriptase and integrase) into the cell.

Reverse Transcription: The single stranded viral RNA is converted into double stranded deoxyribonucleic acid (DNA) using a reverse transcriptase enzyme.

Nuclear Entry and Integration: The viral DNA becomes part of the host genome at a position where transcription takes place regularly. This section of DNA, called a provirus, may stay inactive for a long time (latency). The host cell treats the provirus as part of its own genome and the immune system is unable to detect that the cell is now infected.

Transcription: After latency the viral DNA is activated by changes in the infected cell and starts to produce new virus via transcription (production of ribonucleic acid (RNA)) and translation (production of proteins). This stage uses the body's own replication mechanism.

Viral Assembly and Migration: Protease enzymes cut the HIV proteins into shorter pieces which are transported to the edge of the cell.

Budding: HIV viral particles are covered in the host's cell membrane and are released from the cell completing the process. HIV glycoproteins stick out from the surface of the new virus which helps subsequent binding and fusion. This cycle repeats.

1.2.5 Natural history

1.2.5.1 *Acute Infection*

During acute HIV infection a huge rise in HIV viral load occurs quickly and can reach many millions of viral copies/ml. There is also a temporary drop in circulating CD4 cells numbers. During this period people are much contagious than when the infection becomes chronic [45, 46].

During the acute phase, CD4 memory cells expressing CCR5 co-receptors, are preferentially depleted from gut lymphoid tissue leading to increased gut permeability. This leads to an influx of microbes into the blood stream from the intestine. Inflammatory pathways become upregulated in response [47-49].

1.2.5.2 *HIV seroconversion*

HIV seroconversion is the period early in infection, during which the host forms antibodies to the virus and these antibodies become detectable in bodily fluids. This usually occurs within a few weeks of infection and individuals frequently feel unwell with a, "seroconversion illness". A variety of symptoms commonly occur including: fever, night sweats, tiredness, headaches, sore throat, a rash, weight loss, mouth ulcers, muscle and joint pain and lymphadenopathy [45, 46]. These symptoms are non-specific and during this stage people may either not seek medical care or may not be tested for HIV if they do [2].

1.2.5.3 *Chronic Infection*

After the acute phase, circulating CD4 cells tend to return to near normal levels and viraemia is reduced. HIV viral load is less labile and the viral load "set point" is said to have been established as the host immune system regains some control over viral replication [2]. The level of this "set point" (steady state) is known to be an independent predictor of disease progression [2, 50, 51]. This stage of infection is called the chronic or clinically latent stage and there are usually few symptoms [2]. Although people may feel healthy, much viral replication is still taking place [46]. This replication leads to the progressive loss of circulating T-cell numbers over time, at an average rate of 50-100 cells/µL per year [52, 53].

1.2.6 Co-infection with hepatitis B and C

The prevalence of infection with HCV and hepatitis B (HBV) is much higher in those with HIV than it is in the general population [54, 55]. Liver disease is one of the most common causes of non-AIDS related death in PLHIV and coinfection with hepatitis viruses is present in nearly all these cases [56, 57].

1.2.6.1 *HCV co-infection*

HCV is a major cause of morbidity and mortality world-wide [58, 59]. PLHIV have been found to have six times the odds of being HCV antibody positive (a measure of prior exposure) than those without HIV [54]. HCV coinfection prevalence varies markedly by region within Europe, being lowest in Northern Europe (17.3%) and highest in Eastern and Central Europe (34.0%) [60].

HCV is commonly transmitted through contact with contaminated blood, predominantly through unsafe injections and is more infectious than HIV [59]. It occurs frequently in PLHIV who inject drugs. European studies suggest that in the region of 80-90% of HIV-positive PWID are co-infected with HCV [61, 62]. In contrast, the prevalence amongst MSM in these studies was just 4% [61, 62]. There have been outbreaks of HCV infection in HIV-positive MSM however, probably spread through sexual activity [63, 64]. Those engaging in Chemsex may be at higher risk [63, 64]. In the general population around 20-30% of individuals spontaneously clear HCV infection. A small UK study found that 15% (95% CI, 8-22) of PLHIV were found to do so [65].

The effects of chronic co-infection include progressive liver fibrosis leading to cirrhosis, cancer and renal disease (glomerulonephritis). Infection with HIV markedly increases the rate at which HCV-related liver disease progresses, leading to much poorer outcomes [66]. No vaccine is available for HCV infection, but the development of direct acting antivirals (DAAs) in recent years greatly improved treatment options for those with the virus [59].

1.2.6.2 *HBV co-infection*

Chronic HBV coinfection prevalence amongst PLHIV has been estimated to be about 10% globally and 5%-7% in Western Europe and North America. In these high-income areas the highest prevalence (9%-17%) is found amongst men who have sex with men (MSM) [67].

HBV is transmitted in a similar way to HIV (vertically and through contact with some bodily fluids, especially blood and semen), but is more infectious [68]. Most adults who get HBV will have an effective immune response and develop lifelong immunity. About 10% of adults in the general population and 25% of HIV infected adults however will develop a

chronic infection [67]. Chronic infection frequently leads to liver cirrhosis and liver cancer [68]. The most effective means of prevention is by vaccination prior to exposure. Treatment of chronic infection is recommended, either with interferon or life-long antivirals, but is frequently not very effective [68].

1.2.7 Prevalence and Incidence of HIV

The United Nations (UN) estimates that at the end of 2017 there were about 37 million [95% Confidence intervals (95%CI), 31–44 million] people living with HIV [4]. It is thought that in the region of 35 million [95%CI, 25–50 million] people have ever died from AIDS-related illnesses [4].The HIV prevalence has increased from 28 million [95%CI, 23-32 million] in 2000 (despite declining incidence). This is due to increased survival in individuals accessing treatment. Global incidence peaked in 1997 at about 4 million new cases per year and recent estimates indicate around 2 million new cases occurred in 2017 [4]. The majority of existing and new cases are in Sub-Saharan Africa, where just over half of all HIV-positive people live [4]. Incidence is now declining in Sub-Saharan Africa however, as well as in Western Europe and America [4]. New infections are increasing in Eastern Europe, central Asia, the Middle-East and North Africa where only around a third of PLHIV are receiving treatment [4]. Early treatment of those with HIV is now seen as crucial to controlling the pandemic, because viral suppression is extremely effective in preventing transmission [69, 70].

1.2.8 Treatment of HIV in adults

Several trials of potential HIV vaccines have taken place in the last 25 years but no effective vaccine has yet been developed [71-75]. Research has produced >30 drugs effective in treating HIV in adults [76, 77]. Before 1987 the only treatments were prophylactic antimicrobials, which help prevent opportunistic infections [78]. Treatment with ART started with the introduction of zidovudine (AZT) a nucleoside reverse transcriptase inhibitor (NRTI) in 1987 [79]. This was followed by a period of dual therapy from 1991 with the introduction of didanosine (ddI) (also an NRTI) which could be given in combination with AZT [80, 81]. Trials found lower rates of death with AZT monotherapy in those with AIDS, but no benefit in those without symptoms. Dual therapy was found to modestly slow disease progression and prolong life [82].

In 1996 the first non-nucleoside reverse transcriptase inhibitor (NNRTI) nevirapine [83] and the first protease inhibitor (PI) saquinavir [84] became available. This was the beginning of

the combination ART (cART) era. These treatments led to huge improvements in both life expectancy and quality for PLHIV with access to these drugs [85, 86].

The aim of cART is to inhibit HIV viral replication so the virus becomes undetectable in the blood and CD4 cell counts are maintained or restored to normal levels [78]. Individuals are started on a combination of drugs (usually ≥3) which act at different points in the HIV replication cycle. cART is more effective at suppressing the virus than using one or two drugs because multiple stages of replication are blocked simultaneously [78]. As drugs differ in their ability to penetrate different tissues, combining them also results in effective treatment reaching more of the body [87]. Although cART typically comprises at least three drugs from at least two different classes of ART, recently the two drug combination dolutegravir (DTG) and rilpivirine (RPV) has been licensed [88, 89].

ART classes are defined by the point in the HIV viral replication cycle at which they act [39, 78, 90]:

NRTIs e.g. AZT/ZDV, abacavir (ABC) and emtricitabine (FTC): NRTIs competitively inhibit HIV reverse transcription causing premature termination of the DNA chain.
NNRTIs e.g**.** RPV, efavirenz (EFV) and nevirapine (NVP): HIV reverse transcriptase is made up of two subunits. NNRTIs bind to one of the subunits and reduces its activity.
PIs e.g. indinavir (IDV), lopinavir/ritonavir (LPV/r) and saquinavir: PIs bind to and prevent the protein chains from being cut into shorter polypeptides by the HIV protease enzyme.
Integrase inhibitors (INSTIs) e.g. DTG, raltegravir (RAL) and elvitegravir (EVG): INSTIs inhibit HIV integrase and prevent transport and attachment of viral DNA to the host's DNA.
Fusion inhibitors (FIs)-enfuvirtide (T20): it acts outside the cell by preventing the virus particles from fusing with the cell.
Chemokine receptor antagonists (CCR5 antagonists)-maraviroc (MVC): it acts by binding to the CCR5 on the CD4 cell and prevents fusion.
Recommendations regarding when to start ART, and which combination of drugs to start on, have changed markedly with time as knowledge and treatment options have increased [77, 91-94]. Generally there has been a trend over the years to recommend starting treatment progressively earlier in the disease course [95]. The publication in 2015 of findings from both the Strategic Timing of Antiretroviral Therapy (START) trial (see Section 1.3.5) and the TEMPRANO ANRS 12136 trial led to new recommendations worldwide [91, 96]. START and other studies have shown that early initiation slows disease progression and reduces transmission, morbidity and mortality [92-94, 97-101]. The European AIDS Clinical Society (EACS, Europe), World Health Organisation (WHO, International) [102, 103],

British HIV Association (BHIVA, UK) [104]and Department of Health and Human Services Panel (HHSP, USA) [105, 106] all now recommend that all individuals start cART at diagnosis.

Treatment in low and middle income countries has lagged behind treatment in higher income settings. Although first-line treatment is now more widely available and many country guidelines recommend immediate ART, about 40% of HIV-infected individuals worldwide are still not on treatment [107].

Table 1.1 shows how EACS guidelines of what and when to start have changed over time in Europe since 2005 [95].

Table 1.1

<table>
<tr><th>Year (version)</th><th colspan="3">Recommended initial regimen</th><th colspan="2">When to start treatment in naïve patients with chronic infection</th></tr>
<tr><td rowspan="4">2001-2005</td><td colspan="3" rowspan="4">One of the following regimens:
2NRTI + 1PI (or boosted PI)
2NRTI + 1 NNRTI
3 NRTI</td><td rowspan="3">Recommended</td><td>In all symptomatic patients</td></tr>
<tr><td>CD4<350 cells/µL</td></tr>
<tr><td>CD4 >350 cells/µL & VL >10^5</td></tr>
<tr><td>Consider</td><td>CD4 350-500 cells/µL & VL 50 000-100 000</td></tr>
<tr><td rowspan="10">2005</td><td colspan="3">One drug from each column should be combined</td><td>Mandatory</td><td>In all symptomatic patients</td></tr>
<tr><td>A</td><td>B</td><td>C</td><td rowspan="3">Recommended</td><td>CD4 <200 cells/µL or <15%</td></tr>
<tr><td colspan="3">Recommended</td><td>CD4 201-350 cells/µL and rapid CD4 decline</td></tr>
<tr><td>EFV[1]
NVP[2]
Boosted PI
(FPV[3]/r, LPV[4]/r, SQV[5]/r)</td><td>ABC[6]
TDF[7]
ZDV[8]</td><td>3TC[9]/FTC[10]</td><td rowspan="2">Consider</td><td rowspan="2">CD4 201-350 cells/µL and either:
-HIV viral load >10^5 copies/mL
-HCV co-infection identified
Or if patient seeking treatment and ready to start ART</td></tr>
<tr><td colspan="3">Alternative</td></tr>
<tr><td>ATV[11]/r or IDV[12]/r if above contra-indicated</td><td>ddI[13] if above contra-indicated</td><td></td><td>Defer</td><td>CD4 >350 cells/µL (with close monitoring if viral load >10^5 copies/mL or rapid CD4 decline)</td></tr>
<tr><td rowspan="2">2007 (only changes are listed for this and each subsequent version)</td><td>IDR/r[12] removed</td><td>ZDV[8] moved from recommended to alternative</td><td></td><td>Consider</td><td>CD4 350-500 cells/µL and:
VL>10^5 copies/mL and/or CD4 cell count decline >50-100 cells/µL/year
-age >55
- hepatitis C co-infection</td></tr>
<tr><td>Defer</td><td>CD4 >500 cells/µL (unless patient is seeking treatment and ready to start ART)</td></tr>
</table>

Year (version)	**Recommended initial regimen**			**When to start treatment in naïve patients with chronic infection**	
2008(version 4)	DRV[13]/r added as possible alternative ATV[11]/r moved to recommended				As 2007
2009 (version 5)	DRV[13]/r moved to recommended FPV[3]/r moved to alternative RAL[14] added to alternative			Recommended	CD4 350-500 cells/µL: -HBV con-infection requiring therapy -HIV-associated nephropathy or other organ deficiency
2009 (version 5) continued				Consider	CD4 350-500 cells/µL: -Pregnancy -High cardiovascular (CVD) risk -Malignancy
					CD4 >500 cells/µL and: HCV, HBV requiring therapy; nephropathy or other organ deficiency; pregnancy; high CVD risk or malignancy
2011 (Version 6 & 6.1)	SQV[5]/r moved to alternative Maraviroc (MVC) added to alternative RAL[14] moved to recommended			Consider	CD4 350-500 cells/µL: -Asymptomatic HIV infection -Autoimmune disease -HBV co-infection not requiring treatment
				Recommended	CD4 350-500 cells/µL: -Pregnancy (before 3rd trimester) -High CVD risk -Hodgkin's Lymphoma HPV-associated cancers
2014 (version 7.1)	Rilpivirine (RPV) added to recommended LPV[4]/r and NVP moved to alternative RAL[14] moved to recommended Dolutegravir (DTG) added to recommended Elvitegravir (EVG) + cobicistat (COBI) added to recommended DRV/r[13] + RAL[14] added as an alternative for some patients LPV/r + 3TC added as an alternative			Consider	To reduce HIV transmission in those with CD4 cell counts >350 cells/µL
				Recommended	All pregnancies (before 3rd trimester)

Year (version)	**Recommended initial regimen**	**When to start treatment in naïve patients with chronic infection**	
2015 (version 8)	EFV ATV[11] and some DRV[13] regimens moved to alternative Cobicistat added as an alternative to ritonavir combined with PIs	Strongly recommended	All patients with CD4 count <350 cells/µL
		Recommended	All patients with CD4 count ≥350 cells/µL (with the possible exception of elite controllers)
2015 (version 8.1)	TAF[15] added as an option as part of the recommended NRTI backbone	Recommended	For all adults with chronic HIV infection irrespective of CD4 count
2017 (version 8.2-9.0)	What and when to start recommendations unchanged from version 8.1		
2018 (version 9.1)	Integrase inhibitor (RAL/ DTG/BIC) and two NRTIs recommended: When to start for treatment naïve individuals unchanged		

EFV-Efavirenz [NVP]-Nevirapine [3]Fosamprenavir LPV-Lopinavir [5Saquinavir] [6]Abacavir [7]Tenofovir Disoproxil Fumarate [8]Zidovudine [9]Lamivudine [10]Emtricitabine [11]Atazanavir [12]Indinavir [13]Darunavir [14]Raltegravir [15]Tenfovir Alafenamide RAL-Raltegravir DTG-Dolutegravr BIC-Bictegravir

1.2.9 Survival with HIV

Survival of PLHIV in high-income countries has dramatically increased since the beginning of the epidemic [7]. Recent analysis of data from Europe and North America found that life expectancy increased by over a decade when comparing those starting ART between 2008-2010 with those starting between 1996-1999 [9]. Estimated life expectancy for a European HIV-positive individual starting ART aged 20 years between 2008-2010 was 68 years (95% CI, 67–69) for both men and women. This was lower than in the French general population, where men are expected to live 79 years on average and women 85 years [9]. Other similar studies have showed rises in life expectancy between 15 and 24 years over a similar time period [10, 11, 108]. Whilst survival has greatly improved, it still falls somewhat short of that enjoyed by the general population. The most important predictor of life-expectancy in those with HIV is their level of immune-suppression (low CD4 cell count). Marked immune-suppression is commonly due to HIV diagnosis late in the course of infection, delays in starting cART or poor treatment adherence [7]. PLHIV are also more likely to smoke, inject drugs, abuse alcohol and be co-infected with HBV and HCV than the general population, which are all known to impact life-expectancy [7].

1.3 Serious non-AIDS events

1.3.1 Definition

There is no universally agreed SNAE definition and studies vary with regards to both the conditions included and how each is defined. Previous studies have often examined non-AIDS defining cancers, cardiovascular disease, renal and liver disease [18, 20, 109, 110]. Pancreatitis has sometimes been classed as a SNAE [111] as have psychiatric events [112, 113]. To be defined as “serious” some studies have stipulated that an overnight hospital stay for the event is necessary [113, 114]. One study largely based their definition of SNAEs on Division of AIDS ≥grade 4 adverse events, which also included abnormal laboratory results [115].

The International Network for Strategic Initiatives in Global HIV Trials (INSIGHT) study group formulated robust case definitions for a number of conditions which they classified as SNAEs [19]. These included a number of cardiovascular events: acute MI, congestive heart failure, coronary artery disease necessitating drug treatment, coronary revascularization, deep vein thrombosis, peripheral arterial disease, pulmonary embolism, and stroke. Non-cardiovascular conditions for which they also formulated case definitions

comprised: decompensated liver disease, diabetes mellitus, end-stage renal disease and non-AIDS cancer [19].

1.3.2 Why are SNAEs important?

In resource-rich countries, cardiovascular disease, cancer, liver and kidney disease, DM and fractures are all now leading causes of morbidity and/or mortality, irrespective of HIV status [116, 117].

1.3.3 Evidence of elevated SNAE prevalence and incidence in PLHIV

A number of studies and meta-analyses have found that CVD, non-AIDS defining cancers and fractures occur at higher rates in those with HIV when compared to those without the virus or those in the general population [118-121]. Few studies have compared rates of liver disease [122], kidney disease, diabetes [123, 124] or other non-AIDS conditions in those with HIV to other populations.

Islam et al. undertook a meta-analysis to ascertain the incidence of CVD amongst PLHIV and included results from 20 studies. Compared to HIV negative individuals there was a 60% increase in CVD rate in those with HIV not on ART and the rate doubled in those who were on ART [118]. A recent systematic review and meta-analysis by Rao et al. exclusively examining MI, found that HIV infection was associated with twice the risk of AMI compared to matched HIV-negative controls [125].

Over 400 000 PLHIV from seven studies were included in a meta-analysis comparing the incidence of 28 different types of cancer in those with and without HIV. Most cancers (20/28) occurred at elevated rates in the HIV-positive population [26]. This meta-analysis found that cancers with a known viral aetiology tended to occur at higher rates in the HIV-positive group. [26]. A slightly larger and overlapping meta-analysis reported a Standardised Incidence Ratio (SIR) for all non-AIDS cancers combined of 2.0 (1.8-2.2) [120].

A number of studies have compared fracture rates in HIV-positive and negative individuals, most of which reported fracture rates to be increased in those with HIV [126-132]. A meta-analysis which included data from 13 studies reported an Incidence Rate Ratio (IRR) for fractures (both fragility and high-impact combined) of 1.58 (95%CI, 1.25-2.00) [119].

Studies comparing diabetes mellitus (DM) prevalence and incidence between HIV-positive and negative populations have produced conflicting results [123, 124, 133, 134]. Brar et al., in a US cross-sectional study of 2 565 individuals, did not find those with HIV (who were all ART-naïve) had a higher prevalence of DM compared to those without [133]. Brown et al., however, analysing data on 1 278 individuals in the Multi-centre AIDS Cohort Study

(MACS), found PLHIV (all of whom were on cART) to experience more than four times the prevalence and incidence of DM compared to at-risk HIV negative controls [124]. In a study including all PLHIV treated in Denmark, Rasmussen et al., found DM rates were not elevated in those with HIV when compared to age and sex matched population controls [123]. They did, however, find elevated incidence in the 1996-1999 calendar period for those with HIV, possibly related to ART drugs no longer commonly used [123]. Studies have found liver disease is now the most common cause of non-AIDS death in PLHIV (which is not the case in the European general population where CVD predominates), especially those co-infected with HCV (and to a lesser extent HBV) [135, 136].

1.3.4 Possible reasons for the increased incidence of SNAEs in PLHIV

Although research suggests that some SNAEs occur at higher rates in those with HIV compared to the general population or those without the virus [26-30, 118, 120, 137-139], HIV may not be the cause of the increase. It could be that these findings, all from observational studies could be explained completely or in-part by confounding.

Confounding could occur due to higher rates of exposure to traditional risk factors for SNAEs in those with HIV. Such risk factors include (depending on the condition): smoking, alcohol and drug use, obesity, diabetes, a sedentary lifestyle, high blood pressure and dyslipidaemia [30, 134, 140, 141].

There is increasing evidence that coinfection with other viruses may also impact SNAE risk. HCV, HBC, cytomegalovirus (CMV) and Epstein-Barr virus have all implicated [142-150].

It is also possible that exposure to ART could be the reason for the observed differences in SNAE incidence between HIV-positive and negative populations. CVD, DM and fractures have all been associated with exposure to certain classes of ART or specific drugs [151-154]. For example, the NRTI TDF has been associated with increased fracture incidence and exposure to abacavir and the PI combination lopinavir/ritonavir has been associated with increased rates of MI [154-157]. DM has been associated with exposure to stavudine (d4T), zidovudine (AZT) and didanosine (ddi) and duration of cART [158]. These findings are also from observational studies, however, and so again associations are not necessarily causal [159]. There are particular problems with confounding by indication and time-varying confounding when examining associations between ART exposure and outcomes [160].

1.3.5 Results from clinical trials examining SNAEs

In 2006, The Strategies for Management of Antiretroviral Therapy (SMART) study published their findings [161]. This trial included 5 472 patients who were randomised to one of two arms, the drug conservation (DC) group and the viral suppression (VS) group. In the DC group individuals were given intermittent ART dependent on their CD4 cell count. In the VS group individuals were given continuous ART to suppress HIV viral replication. The trial aimed to determine whether it was better to take ART intermittently as long as the CD4 cell count was maintained above a certain level (≥250 cells/μL in this case), or whether the benefits of ART outweighed its adverse effects and that it would be better to take it continuously. All individuals in the trial had a CD4 cell count >350 cells/μL at baseline. In the DC arm ART was started when the CD4 cell count fell <250cells/μL and was stopped when it rose >350 cells/μL. Deaths from any cause were higher in the DC group compared to the VS group (Hazard Ratio (HR), 1.8;1.2-2.9; p=0.007) and rates of major cardiovascular, renal or hepatic disease (as an aggregate outcome) were also higher in the DC group (HR, 1.7; 1.1-2.5, p=0.009) [162]. There were 104 of these aggregate events in total, with 65 occurring in the DC group and 39 in the VS group. When the outcome was restricted to CVD (both fatal and non-fatal), 9 events occurred in the DC group and 2 in the VS group (HR, 1.6; 1.0-2.5; p=0.05) [161]. Further analysis of the trial data found no difference in rates of non-AIDS defining malignancies (nADM) between the two arms (HR, 1.3; 0.7-2.1; p=0.4). There were 58 of these events in the analysis [163].

This trial results led people to speculate whether uncontrolled HIV infection could be causing these diseases [117, 163]. The findings of SMART led to the design of a further trial, the Strategic Timing of Antiretroviral Treatment (START) trial. This trial randomized 4 685 individuals who had a CD4 cell count >500 cells/μL to start cART immediately or to wait for treatment until their CD4 cell count had fallen ≤350 cells/μL (or the individual had an AIDS-related event/other condition that necessitated starting ART). This trial reported early in 2015 after the start of my PhD, after a mean follow-up time of 3.0 years. A composite serious non-AIDS event outcome which included cardiovascular disease, end-stage renal disease, end-stage liver disease, nADM and other non-AIDS deaths was lower in those who started ART immediately compared to those who deferred (0.61; 0.38–0.97; p=0.04) [91]. There were 76 SNAEs in total, 29 of which were in those who started cART immediately and 47 in those deferring. nADM incidence appeared lower in those immediately starting cART, but was of borderline significance (0.50; 0.22-1.11; p=0.09). No association was found between treatment arm and CVD incidence (0.84; 0.39-1.81; 0.65), but only 26 CVD events

occurred in total [91]. START and SMART did not examine diabetes or fractures, but a START sub-study examined bone mineral density (BMD-a major risk factor for fragility fractures) in 424 ART naïve individuals. It found that both current and nadir CD4 cell count and current HIV viral load were not associated with BMD [164].

The SMART and START findings, suggest that even though exposure to ART may be causally associated with a number of SNAEs, benefits outweigh risks. These findings also suggest that untreated HIV infection could have a role in SNAE incidence.

1.3.6 Other evidence HIV might be causally associated with SNAEs

In HIV-positive people who die, the proportion of deaths attributable to non-AIDS causes have markedly increased with time since 1996 [20, 21, 24, 165]. Some studies have also found that rates of non-AIDS deaths have decreased in the same time period [12, 166]. This led to speculation that HIV could be contributing to non-AIDS deaths and that cART might be protective [117].

Further evidence suggesting that HIV might have a causal role in some nADM is provided by an elegant meta-analysis. This examined the rate of specific nADM in both the HIV-positive population and in recipients of transplants. Cancer rates from both groups were then compared to those of the general population [26]. The rationale for this approach was that PLHIV and transplant recipients are only likely to be similar to each other with respect to their immune-suppression, but are likely to differ with respect to confounders. The authors argued that if the rates of a cancer were elevated in both transplant recipients and PLHIV when compared to the general population, confounding was an unlikely cause. The study reported elevated rates for 20 of the 28 cancers investigated in both patient groups. These 20 mostly comprised cancers with a known infectious cause such as Hodgkin's Lymphoma [26].

So, at the time of starting my PhD project there was evidence from observational studies that some SNAEs occur at higher rates in those with HIV than in the general population. Furthermore, results from a number of clinical trials and observational studies were suggestive that HIV infection might be causally associated with SNAEs.

1.3.7 Possible mechanisms of causal association

If HIV is causally associated with SNAEs, by what mechanisms could this be occurring? Perhaps extended periods of HIV viremia or immuno-suppression predispose individuals to these conditions [117, 167]. Immune suppression leads to reduced immuno-surveillance and/or increased inflammation and immune activation in those with HIV [168, 169]. Even

undetectable HIV viral load and CD4 cell counts within the normal range might increase SNAE risk due to persistent inflammation, endothelial activation and thrombosis [170]. CD4 cells fall to very low levels in gut lymphoid tissue early in HIV infection and do not appear to be restored with effective treatment [171]. The lack of effective immunity within the gut, leads to large numbers of microbes continually entering the blood stream through the ineffective mucosal barrier. This leads to chronic ineffective inflammation, which is known to be detrimental to health [47, 172].

1.3.8 The need for further research

It is not clear what relative contribution HIV infection, ART and traditional risk factors make to elevated SNAE rates. This is likely to vary by SNAE type. The challenge is to try to tease out the contribution of each factor.

One avenue for further research would be to examine whether markers of HIV infection, such as HIV infection duration, HIV viremia and immune-suppression (including current and nadir CD4 cell count and duration of immune-suppression) were associated with the incidence of individual SNAEs. Previous studies have examined these exposures for a number of SNAEs, but most analyses have used data from sero-prevalent cohorts which have incomplete information on markers of HIV [110, 111, 173-179].The aim of my PhD is therefore to examine these markers of HIV infection in the Concerted Action on SeroConversion to AIDS and Death in Europe (CASCADE) cohort collaboration where all individuals have a well-estimated time of seroconversion. I will be able to more accurately estimate duration of HIV infection, nadir CD4 cell count and duration of immune-suppression than is commonly possible in most previous research.

1.4 Thesis outline

1.4.1 Thesis aims

The primary aim of my thesis is to examine the association between duration of HIV infection from a well-estimated time of HIV seroconversion and incidence of individual SNAEs. I will explore these associations both before and after controlling for other factors.

No other studies, to my knowledge, have examined the association between this exposure and SNAE incidence in individuals with well-estimated seroconversion dates.

My secondary aims are to examine how other HIV-related exposures are associated with SNAE incidence, adding to existing research in this area [110, 111, 173-179]. These exposures comprise: current (time-updated) HIV viral load, current CD4 cell count, nadir CD4 cell count, duration of immunosuppression at three levels (≤200/100/50 cells/μL) and prior AIDS.

By examining the associations between HIV-related factors, including duration of HIV infection, and my outcomes I aim to determine (as far as is possible using regression modelling of observational data) the contribution of HIV infection itself to these events.

1.4.2 Thesis structure

Chapter 2: Literature reviews

I describe the findings of three systematic literature reviews of studies examining the association between a number of exposures of interest and fractures, MI and DM. These exposures comprise: duration of HIV infection, CD4 cell count (nadir, current and baseline), HIV viral load (current and baseline) and duration of immunosuppression.

Chapter 3: Survey, case definitions, CASCADE information and SNAE data capture

The first part of this chapter provides details of a survey I undertook of CASCADE cohorts. When I started this PhD, CASCADE was not capturing data on SNAEs and it was not known what data contributing cohorts were collecting. Hence, initial objectives of the project included determining: which events (if any) each cohort was capturing, how individual SNAEs were defined within each cohort, whether cohorts had developed case definitions for SNAEs, how these case definitions were defined and approximately how many SNAEs of each type had been recorded by each cohort. I designed a survey which I sent to cohorts to answer these questions. Details of how I researched and created the survey are provided along with survey results.

I also discuss SNAE case definition formulation and describe the process by which I captured data on SNAEs from collaborating cohorts.

Chapter 4: Methods

I describe the methods for the analysis of fractures, DM and MI and additional sensitivity analyses.

Chapter 5: Patient characteristics

I summarise characteristics of individuals included and excluded from my analyses by describing: inclusion criteria, numbers and reasons for exclusions (both people and events), patient characteristics at various time points, differences in patient characteristics of those included and excluded and the representativeness of my sample to the total HIV-positive population.

Chapters 6-8: Results

In chapters 6-8 I summarise the results of my analyses by outcome. I then go on to compare the characteristics and findings of my study to other relevant studies.

Chapter 9: Discussion

In Chapter 9 I summarise my findings and discuss their implications. I then highlight major study strengths and limitations, evaluate the likely generalisability of my findings to other populations and discuss my results in light of previous research. I go on to assess evidence for causal associations between my exposures and outcomes and finally summarise areas of possible further work.

1.4.3 Work undertaken by myself and the contribution of other people to my project

Professor Cristina Mussini (University of Modena and Reggio Emilia) had the idea of exploring the association between well-estimated HIV infection duration and the incidence of non-AIDS events. This was further expanded by Professor Kholoud Porter and Professor Caroline Sabin of University College London.

CASCADE was capturing data annually at the start of the study. All data routinely collected were cleaned and merged from collaborating cohorts by Ashley Olson, the CASCADE data manager and statistician at that time.

I undertook the literature reviews and created, circulated, queried and collated the survey regarding SNAE data capture by cohorts. The HIV Cohorts Data Exchange Protocol (HICDEP) table which I created and circulated to participating cohorts was based on one developed

and used by EuroSIDA. I liaised with cohort data managers and principal investigators regarding data acquisition. I cleaned the SNAE data and queried cohorts where necessary. All data management needed for the study and all study analysis was undertaken by me using STATA®. All written work, tables, figures and appendices in this thesis are my own work.

In addition to input from my supervisors, I received statistical advice from Dr. Deborah Ford and both she and Professor Ali Judd provided me with feedback on some of my chapters.

2 Chapter 2: Literature Review

2.1 Introduction

I undertook literature reviews to determine what was known about the associations between duration/markers of HIV infection and the following events:

1. Fractures
2. Cardiovascular (CVD) and cerebrovascular disease
3. Diabetes Mellitus (DM)

These reviews aimed to determine the state of current knowledge at the start of my project. I have included relevant research published subsequently at the end of each review.

2.2 Methods

PubMed was used to systematically search for relevant publications. Medical Subject Headings (MeSH terms) used and search dates are outlined in Sections 2.3 to 2.5. Publication inclusion criteria applicable to all reviews are as follows:

Language: English language publications only.

Dates: Only publications made available on or before 31/12/2013 (including e-publications made available ahead of print).

Publication type: Original research.

Study design: Restricted to randomised controlled trials (RCTs), cohort studies and case-control studies.

Participant characteristics: Studies included adults ≥13 years of age. The majority of study subjects were required to be from high-income countries (as defined by the World Bank [180]). This was established by determining study location for all studies (and further exploration for multinational collaborations). My rationale for restriction to high-income countries is that factors and patterns of association may vary markedly between resource-limited and resource-rich environments.

Exposures: Studies included examined at least one of the following:

i. CD4 cell count (baseline, current, nadir or duration of immune-suppression)
ii. HIV viral load (including baseline or current)
iii. Duration of HIV infection (from seroconversion or diagnosis)

Measures of effect: For inclusion, studies were required to have calculated a relevant hazard ratio (HR), incidence rate ratio (IRR) or odds ratio (OR) with 95% confidence intervals (95% CI).

Further event specific criteria are outlined at the beginning of each review.

Meta-analyses, reviews and editorials were excluded, but their references were checked for relevant studies not picked up by the PubMed search.

Abstracts from the International AIDS Society (IAS) Conference on HIV Pathogenesis, Treatment & Prevention, and the International AIDS Conferences were also searched from 2001 until 2013 using the IAS abstract archive [181]. Due to intermittent access to the abstracts from the Conference on Retroviruses and Opportunistic Infections (CROI) prior to 2014 it was not possible to include CROI abstracts.

A flow-diagram outlining exclusions with reasons were generated for each review.

The following details of each selected study were recorded:

a. publication name and year of publication
b. study name and design
c. dates of the follow-up period
d. number of PLHIV included
e. number of events and event type
f. median person-years of follow-up
g. factors the study investigators were able to adjust for.

Resources were not available for the reliability of study selection to be independently assessed by another reviewer.

2.3 Fractures

2.3.1 Introduction

This literature review examined the evidence for an association between duration/markers of HIV infection and fractures.

2.3.2 Methods

Inclusion criteria (in addition to those already outlined in 2.2) are as follows:

Outcome: fragility fractures (however defined) or all fracture types combined.

Studies restricted to high-impact fractures were theoretically excluded (but none were found).

PubMed® advanced search: ("hiv"[MeSH Terms] OR "hiv"[All Fields]) AND (("fractures, bone"[MeSH Terms] OR ("fractures"[All Fields] AND "bone"[All Fields]) OR "bone fractures"[All Fields] OR "fractures"[All Fields]) OR osteopaenia[All Fields] OR ("bone diseases, metabolic"[MeSH Terms] OR ("bone"[All Fields] AND "diseases"[All Fields] AND "metabolic"[All Fields]) OR "metabolic bone diseases"[All Fields] OR "osteopenia"[All Fields]) OR ("osteoporosis, postmenopausal"[MeSH Terms] OR ("osteoporosis"[All Fields] AND "postmenopausal"[All Fields]) OR "postmenopausal osteoporosis"[All Fields] OR "osteoporosis"[All Fields] OR "osteoporosis"[MeSH Terms]) OR ("bone density"[MeSH Terms] OR ("bone"[All Fields] AND "density"[All Fields]) OR "bone density"[All Fields] OR ("bone"[All Fields] AND "mineral"[All Fields] AND "density"[All Fields]) OR "bone mineral density"[All Fields])).

Bone mineral density (BMD), osteopaenia and osteoporosis were included in the search terms, despite not being outcomes of interest, because their inclusion improved capture of publications where fractures were an additional outcome.

In addition, the IAS abstract archives were searched from 2001-2013 using the search term “fractures” in the search box.

2.3.3 Results

2.3.3.1 *Study Characteristics*

The search process identified 907 publications of which 43 were identified as potentially relevant and their full text was reviewed. After applying all selection criteria nine publications reporting on data from eight studies were included in the review. Eight of these were reported in journal articles and one in a conference abstract. Table 2.1a provides information on each study. Data from one study, the HIV Outpatient Study (HOPS), were used in two publications. The first study only used HOPS data [182] and the second study combined HOPS data with data from the Study to Understand the Natural History of HIV/AIDS in the Era of Effective Therapy (SUN) [183]. No double-reporting occurred due to no overlap in follow-up dates with respect to HOPS.

Articles were published between 2010 and 2013. Seven studies were cohort studies and one study was case-control. Five studies reported fractures of all types and three studies were restricted to fragility fractures. Fragility fracture definitions varied between studies; one study (two publications) identified fragility fractures by location [132, 184] and the other two by the energy of the impact [178, 185]. Five studies were undertaken in the US and one each from Denmark, Australia and Switzerland.

For the longitudinal studies, median person-years of follow-up (PFYU) ranged from 2.6-7.3 years and the number of fractures included from 37-806. Figure 2.1 shows the search process and study selection.

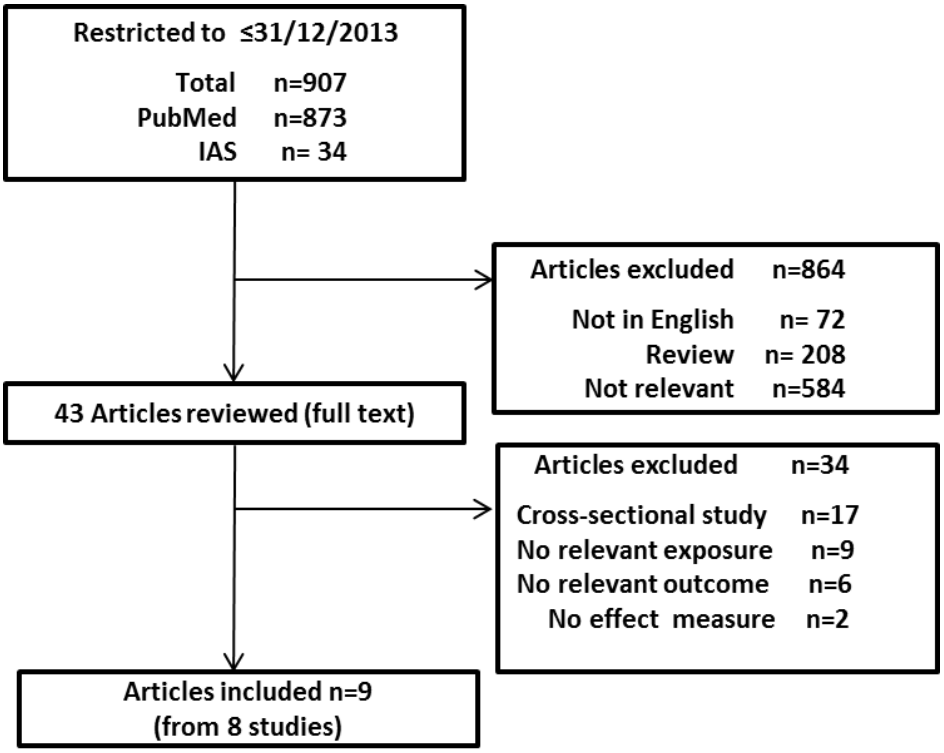


Figure 2.1: Study selection flow diagram: publications examining the association between markers of HIV infection and fractures ≤31/12/2013

Table 2.1a summarises characteristics of studies which examined associations between HIV-related factors of interest to me and fractures.

Table 2.1a: Characteristics of studies examining the association between markers of HIV infection and fractures

Study characteristics								Potential confounders controlled for in the analysis*												Exposure	
Study	First author, year, [paper]	Fracture type	Follow-up period	Source of individuals	Number HIV positive (N)	Number of fractures	PYFU (median)	cART use	Ethnicity	Prior AIDS	Mode of Infection	Smoking	BMI	Steroid use	Secondary Osteoporosis/ Charleston score	HCV	Lipodystrophy	Alcohol	Proton pump inhibitors	CD4 cell count	HIV Viral Load
SUN/HOPS (US)	Battalora 2013 [183]	All	2004-2012	C	1 008	95	5.0	x	√	x	√	√	x	x	x	√	x	x	x	√	x
SHCS (Switzerland)	Hasse 2011 [185]	Fragility	2008-2010	C	8 444	37	2.6~	x	√	X	√	√	x	x	x	x	x	x	x	√	√
DHCS (Denmark)	Hansen 2012 [131]	All	1995-2010	C	5 306	806	6.5	x	√	√	x	√	x	x	√	x	x	x	x	√	x
VACS-VC (US)	Womack 2011 [132]	Fragility	1997-2009	C	40 115	602	6.0~	√	√	x	x	√	√	√	√	√	x	√	√	√	x
	Womack 2013 [184]	Fragility	1997-2009	C	40 115	558	6.0~	√	√	x	x	√	√	√	√	√	x	√	√	√	√
WIHS (US)	Yin 2010 [186]	All	2002-2008	C	1 728	148	5.4	√	√	x	√	√	√	x	√	√	x	√	x	√	x
ALLRT (US)	Yin 2012 [187]	All#	2009-2011	C	4 640	106	5.0	√	√	√	√	√		√	x	√	x	x	x	√	√
Alfred Hospital, (Australia)	Yong 2011 [178]	Fragility	1998-2009	CC	183	73	¥	x	√	x	√	x	√	x	x	√	x	x	x	√	√
HOPS (US)	Young 2011 [182]	All	2000-2008	C	5 826	233	3.8	√	√	√	x	√	√	x	√	√	√	√	√	√	√

*All adjusted for sex and age. ~Mean rather than median. CC-Case-control study C-cohort study U- Unknown BMI-Body Mass Index ICD-International classification of diseases HOPS-HIV Outpatient Study SUN-The Study to Understand the Natural History of HIV/AIDS in the Era of Effective Therapy SHCS-Swiss Hospital Cohort Study DHCS-Danish HIV Cohort Study VACS-VC-Veterans ageing cohort study-virtual cohort WIHS-Women's interagency HIV study ¥-PYFU not applicable as the study was a case-control study #Fractures of the face, skull and digits were excluded

2.3.3.2 *The association between duration of HIV infection and fractures*

Only the Swiss HIV Cohort Study (SHCS) examined the association between duration of HIV infection and fractures (Table 2.1a). The variable was defined as time since the first known positive HIV test at baseline. No point estimate or 95%CIs were reported. This was due to a lack of evidence for an association with fractures ($p \geq 0.05$).

2.3.3.3 *The association between current CD4 cell count and fractures*

The SHCS examined the association between time-updated CD4 cell count and fragility fracture incidence. They found that the CD4 cell count was inversely associated with fracture risk (HR for square root CD4, 0.90; 95%CI, 0.85–0.95, $p<0.001$) after adjustment. The analysis adjusted for $\log_{10}$ HIV viral load, sex, PWID status and smoking, but did not adjust for ART exposure, HCV or other established (FRAX) risk factors [185].

In an analysis of data from the US Veterans Ageing Cohort Study-virtual cohort (VACS-VC), Womack et al. examined the association between time-updated CD4 cell count and fragility fractures. No evidence of an independent association was found between current CD4 cell count and fracture incidence (HR per 100 cell/µl decrease, 0.99; 0.95-1.02). The analysis adjusted for: age, sex, ethnicity, alcohol, steroid, smoking, Body Mass Index (BMI), major depressive disorder, stroke and/or cerebrovascular disease, coronary artery disease and/or DM, haemoglobin, fibrosis-4 index, HCV, estimated glomerular filtration rate (eGFR) and current use of TDF, PI and efavirenz [184].

2.3.3.4 *The association between baseline CD4 cell count and fractures*

Six studies examined the association between baseline (variously defined, see below) CD4 cell count and fracture incidence and only one found statistical evidence for an association. An Australian case-control study, undertaken by Yong et al., compared index CD4 cell counts (how index was defined was not reported). Individuals were age, sex and HIV infection duration matched to controls without fractures. In a univariable analysis those with CD4 cell counts <200 cells/µL (OR, 6.77; 2.40-19.10; $p<0.01$) and 200-500 cells/µL (OR, 2.40; 1.06-5.44; $p=0.04$) had higher odds of fragility fracture when compared to those with CD4 >500 cells/µL (reference group). In multivariable analysis those with CD4 cell counts <200 cells/µL remained at elevated odds of fracture (OR, 4.91; 1.78-13.57; $p=0.002$), but it was unclear what had been adjusted for and whether the reference group was those with a CD4 cell count >500 cells/µL or whether the variable was then treated as binary with the reference group having a CD4 cell count ≥200 cells/µL. The multivariable adjustment may have included exposure to steroids and anti-epileptic medication [178].

A US study, undertaken by Battalora et al., combined data from the HIV Outpatient Study (HOPS) and The Study to Understand the Natural History of HIV/AIDS in the Era of Effective Therapy (SUN) [183] (Table 2.1a). In univariable analysis no association was found between baseline (date of first dual energy X-ray absorptiometry (DEXA) scan) CD4 cell count and fracture incidence (HR per 100 cell/µL decrease, 1.04; 0.97-1.11). This variable was therefore excluded from the multivariable analysis.

In a large Danish study, Hansen et al. examined the incidence of low-impact fractures in all HIV-positive HCV-negative individuals receiving care at HIV centres in Denmark. CD4 cell counts at cART initiation were examined. No association was found between baseline CD4 counts >200 cells/µL (IRR, 0.97; 0.73-1.29) when compared to those with CD4 counts ≤200 cell/µL without adjustment. A similar lack of evidence for an independent association with fractures was observed (1.02; 0.75–1.38). Adjustment comprised: age, sex, ethnicity, Charlston's comorbidity index and prior AIDS [2].

The VACS analysis (Table 2.1a) also examined the association between baseline (cohort enrolment) CD4 cell count and fragility fractures. No independent association was found between CD4 cell count at study entry and fracture incidence (HR per 100 cell/µl decrease, 1.01; 0.99-1.70). Adjustment was undertaken for: age, sex, ethnicity, alcohol use, liver disease, steroids, smoking, BMI, lung disease, peripheral vascular disease, major depression, coronary artery disease, DM, congestive heart failure, renal insufficiency and current use of TDF and PI [132].

Yin et al., for the US Women's interagency HIV study (WIHS), examined baseline (cohort enrolment) CD4 cell count in univariable analysis and found no association with fractures (HR per 50 cell/µL increase, 0.97; 0.91-1.03; P=0.3). This study identified fractures through self-report. As no association was found in univariable analysis, baseline CD4 cell count was not included in the multivariable analysis [186].

In the HOPS only analysis, Young et al. found no association between baseline CD4 cell count and self-reported fractures in univariable analysis. Baseline was defined as 1/1/2002 or first HOPS visit thereafter. Those with CD4 cell counts 200-349 cells/µL (HR, 1.07; 95%CI, 0.74-1.55; p=0.7) and those with counts <200 cells/µl (1.28; 95%CI, 0.88-1.86; p=0.2) were compared to those with counts ≥350 cells/µl. Baseline CD4 cell count was excluded from the multivariable model due to lack of evidence for an association[9].

Examining data from the AIDS Clinical Trials Group (ACTG) Longitudinal-Linked Randomized Trial (ALLRT) Yin et al. found no evidence of an association between baseline (at parent

study entry) CD4 cell count and fracture incidence in univariable analysis (Hazard Ratio (HR) per 50 cell/µL increase, 0.98; 0.93-1.03; p=0.4) and so the variable was not included in the multivariable model [187].

2.3.3.5 *The association between nadir CD4 cell count and fractures*

Four studies examined the association between nadir CD4 cell count and fractures, with conflicting results.

In the SUN/HOP analysis, Battalora et al., found no evidence of an association between nadir CD4 cell count and incident fractures in a univariable analysis (HR per 100 cell/µl increase, 1.00; 0.88-1.14) and therefore was not included in the multivariable analysis [183].

In the larger study using HOPS data (233 fractures included vs. 95 in the SUN/HOPS study), Young et al. examined the association between nadir CD4 cell count and fracture incidence in a multivariable analysis and found that those with very low nadir CD4 cell counts experienced higher rates of fracture. Individuals with CD4 cell counts 200-349 cells/µl (HR, 1.25; 0.80-1.92; p=0.3) and with counts <200 cells/µl (1.60; 1.11-2.31; p=0.01) were compared to those with counts ≥350 cells/µL. The model was adjusted for smoking, alcohol use, cART and HCV amongst other variables [182].

Neither WIHS (HR per 100 cell decrease, 0.96; 0.87-1.06; p=0.4) [186] nor ALLRT (HR per 50 cell/µL increase, 0.98; 0.93-1.04; p=0.6 [187] found evidence of an association between nadir CD4 cell count and fracture incidence in univariable analysis and both did not consider the variable further.

2.3.3.6 *The association between HIV viral load and fractures*

There were four studies which examined the association between HIV viral load at baseline and fractures, none of which found evidence of an association.

The Australian case-control study, undertaken by Yong et al., compared fragility fracture odds in those with HIV viral loads >400 copies/mL to those with ≤400 copies/mL (at index, but the time-point at which index occurred was not reported). They found no evidence for an association in univariable analysis (OR, 1.69; 95%CI, 0.97-1.32; p = 0.2) and viral load was dropped [178]. However, the confidence intervals or the OR reported are incorrect (in Table 1 of their paper), because the OR is not included in the CI range.

The HOPS study, compared rates of fractures in those with baseline (1/1/2002 or first HOPS visit thereafter) HIV viral loads >400 copies/mL to those <400 copies/mL. They found no evidence of an association (HR, 0.97; 0.73-1.29; p=0.9) in univariable analysis [182].

In the VACS analysis, Womack et al. examined the association between baseline (at cohort enrolment) HIV viral load and fragility fractures. Viral load was excluded from the final model due to reported collinearity with CD4 cell count (although the reported correlation coefficient of 0.3 was weak). When it was included in the model without CD4, the authors reported that there was no evidence of an association with the outcome, but no measure of its effect was presented [132]. In a subsequent analysis, time-updated HIV viral load was found to be associated with fractures (HR per $\log_{10}$ decrease, 0.91; 0.88-0.94) after adjusted for multiple factors previously described for current CD4 cell count above [184].

In the ALLRT analysis, no evidence of an association between baseline (at parent study enrolment) $\log_{10}$ HIV viral load and fracture incidence was found in the univariable analysis (HR, 0.98; 0.93-1.04; p=0.6) and so it was not included in the multivariable model [187].

2.3.4 The selection of *a priori* confounders

There are a number of well-established risk factors for fractures in those with HIV. Some are HIV specific (e.g. ART) and others (e.g. age) are shared with the general population. These variables have the potential to act as confounders in my analysis [131, 132, 154, 186-194], although factors associated with my exposures of interest are not well-established.

Of these risk factors, CASCADE captured data in 2013 on: age, sex, ethnicity, ART (including exposure to TDF and PIs) and HCV-seropositivity. I included these variables *a priori* in my multivariable models regardless of their statistical significance.

2.3.5 Recent Literature

Since the beginning of 2014 when I undertook my literature reviews, there have been a number of relevant publications.

The Multicentre AIDS Cohort Study (MACS) [195], examined the association between both index (01/01/2001) CD4 cell count (<500 vs. ≥500 cells/μL) and index HIV viral load (≥400 vs. <400 copies/mL) and fractures (Table 2.1b). For all fracture types combined (IRR, 1.22; 95%CI, 0.88-1.69; p=0.2) and for fragility fractures only (IRR, 0.94; 0.54-1.65; p=0.8) no evidence of an independent association between CD4 and fractures was found. Evidence for an independent association between HIV viral load and both fractures of all types (IRR, 0.73; 0.45-1.17; p=0.2) and fragility fractures alone (IRR, 0.76; 0.35-1.68; p=0.5) was similarly lacking.

The WIHS analysis undertaken by Yin et al. (Section 2.3.3) was updated by Sharma in 2015 [128, 196]. The updated analysis included time-updated CD4 cell count and CD4 nadir as opposed to index values which were used in the original analysis. In the univariable analysis there was no evidence that current CD4 cell count (HR, 0.97; 0.93-1.01; p=0.1) was associated with fractures. This variable was not considered further. Evidence of an association between nadir CD4 cell count and fractures was evident however (HR, 0.91; 0.84-0.98; p=0.008). During multivariable model building nadir CD4 was dropped (p>0.05) and adjusted values were not reported.

A US study reported by Gedmintas et al. undertaken at Boston Hospitals included nadir CD4 cell count and prior AIDS as covariates in the analysis [197]. In univariable analysis both nadir CD4 cell count <200 vs. ≥200 cells/μL (RR, 3.7; 2.3–5.8; p <0.01) and prior AIDS (RR, 2.1; 1.5–2.8; p <0.01) were associated with fractures. The point estimates and strength of associations remained similar after adjustment for both nadir CD4 (RR, 3.1; 1.9-5.0; p<0.01) and prior AIDS (RR, 1.6; 1.1–2.2 <0.01).

My original literature review included an abstract by Battalora et. al [183]. This analysis subsequently led to a recent (2016) publication [198]. This slightly extended analysis examined the association between nadir CD4 cell count and fracture incidence, but no evidence for an association was evident either in the univariable (HR, 0.97; 95%CI, 0.89-1.05; p=0.5) or multivariable (0.96; 0.87-1.06; 0.4) results.

Table 2.1b: New literature examining fracture incidence (fragility or all-cause) published since my literature review (01/01/2014)

Study characteristics								Potential confounders on which data were captured*												Exposure		
Study	First author/year [paper]	Fracture type	Follow-up period	Source (individuals)	Number HIV +ve	N (fractures) Fragility/All in HIV+	Total PYFU (median)	cART use	Prior AIDS	Mode	Smoking	BMI	Steroid use	HCV	Opiates	Hypertension	Alcohol	Renal insufficiency	Diabetes	CD4 count	Nadir CD4 count	Viral Load	Prior AIDS
MACS[1] (US)	Gonciulea 2017 [195]	Fragility/ All	2004 - 2012	C	1 221	70/ 182	36 050	√	√	x	√	√	x	√	x	√	√	√	√	√[2]	x	√[2]	√
WHIS[3] (US)	Sharma 2015 [196]	Fragility/ All	2002 - 2013	C	1 713	82/ 300	(10)	√	√	√[4]	√	√	x	√	√	x	√	√	x	√[2]	√	x	√
Boston (US)	Gedmintas 2017 [197]	All	2001 - 2011	C	2 663	180	8 269	√	√	√[4]	√	√	√	√	x	x	√	√	√	x	√	x	√
HOPS/SUN (US)	Battalora 2016 [198]	All	2004 - 2012	C	1 006	85	4 068	√	√	√[4]	√	√	√	√	x	x	√	√	√	x	√	x	√

C-Cohort *All adjusted (or restricted) for age, sex and ethnicity [1]MACS-Multicentre AIDS Cohort Study (men only) [2]Time-fixed value at index (defined as 1st visit after self-reported fracture data capture started in MACS in those ≥40 years of age with ≥1 subsequent follow-up visit) [3] WHIS-Women HIV Interagency Study (also adjusted for menopause, HRT, HBV, prior fracture and use of statins, Vitamin D and calcium); this additional publication updated study findings using more follow-up to augment the publication by Yin in 2010 [186]. [4] Adjusted for injected drug use only

2.3.6 Summary

No studies reported a measure of effect for duration of HIV infection as an exposure. SHCS did examine it however and found no evidence of an association with fractures.

I identified three studies which examined the association between time-updated CD4 cell count and fracture incidence with conflicting results. Elevated rates of fractures in those with lower current CD4 cell counts were found in SHCS but not VACS and WIHS. With one exception (the only case-control study) no evidence was found of an association between baseline CD4 cell count and fractures. Evidence of an association between baseline HIV viral load and fractures was also lacking across all studies.

The one study to examined time-updated HIV viral load found a positive association after adjustment. Results from studies examining nadir CD4 cell count were conflicting, with two US studies finding an association with very low nadir CD4 cell counts and fractures whilst the other three studies did not. No studies examined duration of immune-suppression as an exposure.

This literature review highlights the paucity of research in this area, the conflicting findings of the few studies that have been undertaken, and the need for further research.

2.4 Cardiovascular Disease

2.4.1 Introduction

This literature review examines the evidence for an association between duration/markers of HIV infection and cardiovascular disease (CVD).

2.4.2 Methods

Specific inclusion criteria, in addition to those already outlined in Section 2.2 are as follows:

Outcomes: CVD, ischemic heart disease (IHD), coronary artery disease (CAD), myocardial infarction (MI), stroke (CVE), transient ischemic attack (TIA) and peripheral vascular disease.

Some of these outcomes are a sub-set of each other (e.g. MI and stroke are components of CVD) or different names for the same outcome (e.g. IHD and CAD).

Sub-clinical disease and markers of inflammation and endothelial activation (such as C-reactive protein and D-dimer) were not included.

PubMed® advanced search: (("hiv"[MeSH Terms] OR "hiv"[All Fields]) AND ("cardiovascular diseases"[MeSH Terms] OR ("cardiovascular"[All Fields] AND "diseases"[All Fields]) OR "cardiovascular diseases"[All Fields] OR ("cardiovascular"[All Fields] AND "disease"[All Fields]) OR "cardiovascular disease"[All Fields])) OR (("hiv"[MeSH Terms] OR "hiv"[All Fields]) AND ("ischaemic heart disease"[All Fields] OR "myocardial ischemia"[MeSH Terms] OR ("myocardial"[All Fields] AND "ischemia"[All Fields]) OR "myocardial ischemia"[All Fields] OR ("ischemic"[All Fields] AND "heart"[All Fields] AND "disease"[All Fields]) OR "ischemic heart disease"[All Fields] OR "coronary artery disease"[MeSH Terms] OR ("coronary"[All Fields] AND "artery"[All Fields] AND "disease"[All Fields]) OR "coronary artery disease"[All Fields] OR ("ischemic"[All Fields] AND "heart"[All Fields] AND "disease"[All Fields]))) OR (("hiv"[MeSH Terms] OR "hiv"[All Fields]) AND ("myocardial infarction"[MeSH Terms] OR ("myocardial"[All Fields] AND "infarction"[All Fields]) OR "myocardial infarction"[All Fields])) OR (("hiv"[MeSH Terms] OR "hiv"[All Fields]) AND ("cerebrovascular disorders"[MeSH Terms] OR ("cerebrovascular"[All Fields] AND "disorders"[All Fields]) OR "cerebrovascular disorders"[All Fields] OR ("cerebrovascular"[All Fields] AND "disease"[All Fields]) OR "cerebrovascular disease"[All Fields])) OR (("hiv"[MeSH Terms] OR "hiv"[All Fields]) AND ("stroke"[MeSH Terms] OR "stroke"[All Fields])) OR (("hiv"[MeSH Terms] OR "hiv"[All Fields]) AND ("transient ischaemic attack"[All Fields] OR "ischemic attack, transient"[MeSH Terms] OR ("ischemic"[All Fields] AND "attack"[All Fields] AND "transient"[All Fields]) OR "transient ischemic attack"[All Fields] OR

("transient"[All Fields] AND "ischemic"[All Fields] AND "attack"[All Fields])) OR (("hiv"[MeSH Terms] OR "hiv"[All Fields]) AND ("peripheral vascular diseases"[MeSH Terms] OR ("peripheral"[All Fields] AND "vascular"[All Fields] AND "diseases"[All Fields]) OR "peripheral vascular diseases"[All Fields] OR ("peripheral"[All Fields] AND "vascular"[All Fields] AND "disease"[All Fields]) OR "peripheral vascular disease"[All Fields])) OR (("hiv"[MeSH Terms] OR "hiv"[All Fields]) AND non-AIDS[All Fields]).

Abstracts from the International AIDS Society (IAS) and AIDS conferences were also searched 2001-2013 using the search term “cardiovascular disease” and “cerebrovascular disease” in the conference abstracts search box.

Figure 2.2 shows the numbers of publications identified and excluded with reasons for exclusion and the final number of studies included.

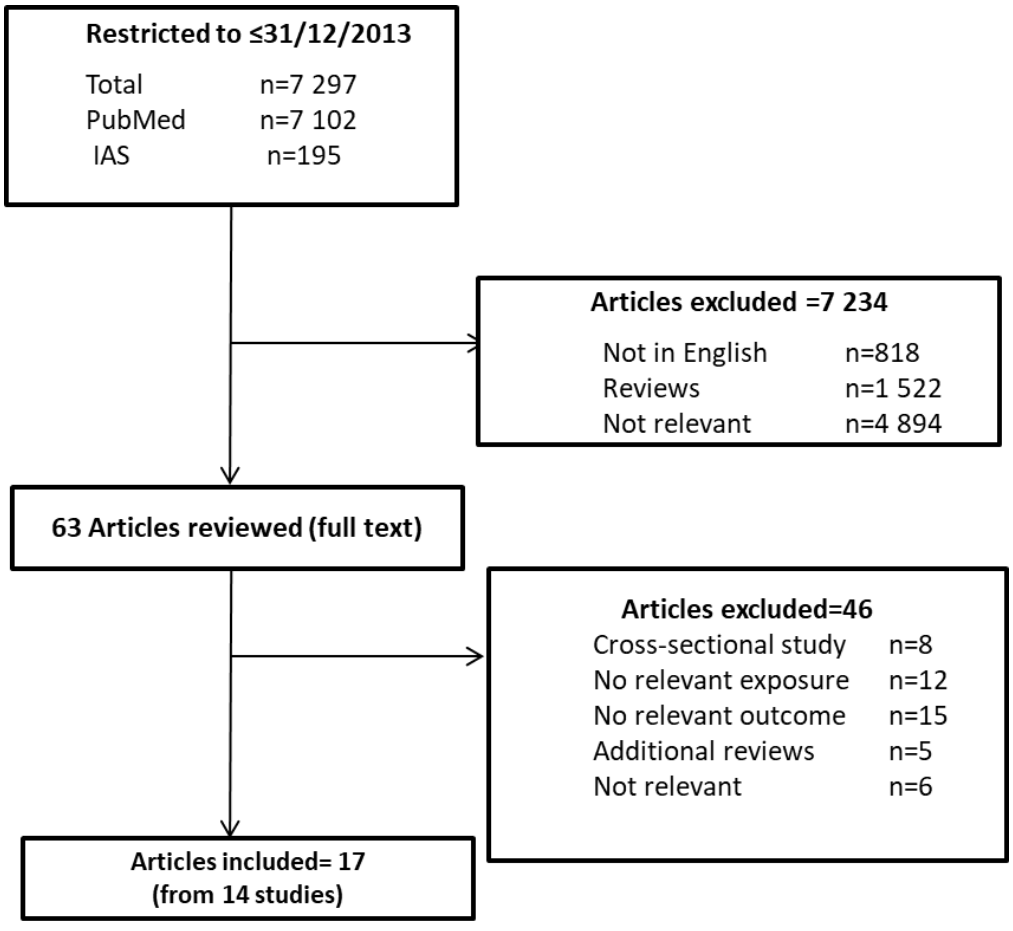


Figure 2.2: CVD study selection flow diagram: publications examining the association between markers of HIV infection and MI ≤31/12/2013

2.4.3 Results

2.4.3.1 *Study Characteristics*

A total of 7 297 publications were found (Figure 2.2) of which 63 were identified as potentially relevant and their full text was reviewed. The inclusion criteria were met by 17 publications from 14 studies. All publications identified were available as journal articles, as opposed to solely as abstracts. Fourteen publications undertook cohort analyses, two reported the results of nested case-control studies and one publication combined both a cohort and case-control analysis [175, 199, 200]. Of those reporting on cohort studies, three publications used data captured during clinical trials [110, 176, 201]. The outcome(s) varied by publication: three examined MI only [199, 202, 203], one coronary artery or other arterial disease events (CADE) only [204], two CVE [205, 206] and the others examined CVD +/- its components. Publication dates ranged from 2007-2013. Four studies were from the US [110, 175, 203, 205, 207], two from France [199, 204], one from Denmark [206, 208], one from Holland [209]. The remaining five were international collaborations [21, 176, 201, 202, 210, 211].

Details of studies included and their relevant publications are shown in Table 2.2a.

Table 2.2a: Characteristics of studies examining the association between markers of HIV infection and CVD/CVE (continues)

	Study characteristics							Data captured on potential confounders*													Exposure		
Study	First author, year, [paper]	Outcome	Follow-up period	HIV positive individuals (N)	Events in PLHIV (n)	Morbidity or mortality	PYFU (median)	Ethnicity	Mode of Infection	CVD/CVE in family	Prior CVD/CVE	Smoking	BMI	cART exposure	Diabetes Mellitus	Renal failure	Hypertension	Lipids	Alcohol	Cocaine	HIV Viral Load	CD4 cell count	Duration of HIV infection
APROCO-COPILOTE (France)	Carrieri 2012 [204]	CADE	1997-2008	1 154	49	Both	U	x	x	√	U	√	x	√	x	x	x	x	√	x	√	√	√
ART-CC (International)	ART-CC 2010 [21]	CVD	1996-2006	39 272	126	Both	(3.6)	x	√	x	x	x	x	√	x	x	x	x	x	x	√	√	x
ATHENA (Holland)	Van Lelyveld 2012 [209]	CVD	2000-2009	3 068	57	Both	10 956	√	√	√	√	√	x	√	x	x	x	x	√	x	x	√	x
CASCADE (International)	Marin 2009 [210]	CVD	1996-2006	9 858	36	Death	71 230	x	√	x	x	x	x	√	x	x	x	x	x	x	√	√	x
D:A:D (International)	Friis-Moller 2007 [202]	MI	1999-2005	23 437	345	Both	94 469	**√**	**√**	√	√	√	√	**√**	√	x	√	√	x	x	x	√	x
	Sabin 2013 [211]	CVD¥	1999-2011	33 301	1284	Both	221505	√	√	√	√	√	√	√	√	x	√	√	x	x	x	√	x
DHCS (Denmark)	Helleberg 2013 [208]	CVD	1995-2010	2 584	138	Both	(4.7)	√	√	x	√	x	x	√	√	x	x	x	x	x	x	√	x
	Rasmussen 2011 [206]	CVE	1995-2010	5 031	140	Both	U	√	√	√	√	x	x	√	x	x	x	x	√	√	x	√	x
ESPRIT/SILCAAT (International)	Acchra 2010 [176]	CVD	2000-2009[1]	3 012	125	Both	(7)	√	x	x	x	x	x	√	x	x	x	x	x	x	x	√	x

All studies were cohort CVD-Cardiovascular disease CADE- coronary artery or other arterial disease events MI-Myocardial Infarction CVE-cerebrovascular events U-unknown C-cohort PLHIV-people living with HIV B-morbidity T-mortality ART-CC CASCADE-Concerted Action on Seroconversion to AIDS and Death in Europe D:A:D-Data collection on Adverse events of Anti-HIV Drugs DHCS-Danish HIV Cohort Study [1]RCT data were used for a cohort analysis *All studies adjusted for sex and age ¥Also included CHD, stroke and MI as outcomes

Table 2.2a: Characteristics of studies examining the association between markers of HIV infection and CVD/CVE (continued)

Study characteristics								Data captured on potential confounders*												Exposure		
Study	First author, year, [paper]	Outcome	Follow-up period/ study type	HIV positive individuals (N)	Events in PLHIV	Morbidity or mortality	PYFU (median)	Ethnicity	Mode of Infection	CVD/CVE in family	Prior CVD/CVE	Smoking	BMI	cART exposure	Diabetes Mellitus	Renal failure	Hypertension	Lipids	Alcohol	Cocaine	HIV viral load	CD4 cell count	Duration of HIV infection
FIRST (US)	Baker 2008 [110]	CVD	1999-2005 Cohort[2]	1397	24	Both	(5.0)	√	√	√	x	x	x	√	x	x	x	x	x	x	x	√	x
FHDH (France)	Lang 2012 [199]	MI	Case-control	1173	289	Both	n/a	x	x	√	-	√	√	√	√	x	√	√	x	√	√	√	x
HOPS (USA)	Lichtenstein 2010 [175]	CVD	2002-2009 Case-Control[3]	2005	148	Both	n/a	√	√	√	√	√	x	√	√	x	√	√	√	x	√	√	x
MGH/BWH (USA)	Triant 2010 [203]	MI	1998-2008 Cohort	6 517	273	Both	-	√	x	x	√	√	x	√	√	x	√	√	x	x	√	√	x
	Chow 2012 [205]	CVE	1996-2009 Cohort	4 308	132	Both	25100	√	x	x	√	√	x	√	√	x	√	√	x	x	√	√	x
SMART (International)	Philips 2008 [201]	CVD	2002-2006 RCT[1]	5 472	79	Both	(2.0)	√	√	x	√	≠	√	√	√	x	√	√	x	x	√	x	x
SHCS (Switzerland)	Bucher 2012	CHD	Case-Control	490	98	Both	n/a	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	x
VACS-VC (USA)	Freiberg 2013 [207]	CVD	2003-2010	27 350	508	Both	(5.9)	√	x	x	x	√	√	x	√	√	√	√	√	√	√	√	x

*All studies adjusted for sex and age B-morbidity. T-mortality C-Cohort CC-case-control study R-randomised controlled trial CVD-Cardiovascular disease CVE-Cerebrovascular event MI-Myocardial infarction CHD-Coronary heart disease SMART- Strategies for Management of Antiretroviral Therapy VACS-VC- Veterans Aging Cohort Study PYFU-person-years of follow-up BMI-Body Mass Index [1]Post-hoc analysis of data from an RCT [2] RCT data used for cohort analysis [3] Both a cohort and case-control analysis were included in the publication ≠cases matched for smoking status with controls

2.4.3.1.1 The association between known HIV infection duration and CVD

Only the French (ANRS CO8) APROCO-COPILOTE cohort study reported a measure of effect for known HIV infection duration. Carrieri et al. undertook the analysis exclusively in PI-exposed PLHIV. There was no evidence that time since HIV diagnosis at cohort enrolment was associated with increased incidence of CADE in univariable analysis (HR per additional year, 1.01; 0.94-1.08; p=0.8) and the variable was therefore not included in the multivariable model [204].

2.4.3.1.2 The association between CD4 counts and CVD

Studies have examined various CD4 cell count metrics including: baseline (usually defined as enrolment or at ART initiation), current, nadir, slope and duration of immune suppression. There is much study heterogeneity with respect to: outcomes definition, CD4 measures examined, exposure categorisation and adjustment.

The association between CD4 cell count at baseline (ART initiation) and CVD death was examined in from 13 cohorts in ART-CC [21]. CVD was assigned as the cause of death in 103 cases. No evidence of an association was found between baseline CD4 cell count and CVD death (HR per 100 cell/µL decrease, 1.08; 0.95-1.23) in multivariable analysis. Adjustment was undertaken for: age, sex, PWID status, HIV viral load (at ART start), prior AIDS, cohort and year of ART initiation [21].

An analysis by van Lelyveld et al. for the Dutch ATHENA cohort took a slightly different approach. They examined how lack of immune recovery despite successful viral suppression after starting cART was associated with CVD incidence [209]. CVD was defined as coronary procedures, MI and stroke. The analysis divided individuals into four groups based on their CD4 cell count two years after starting ART (<200, 200-350, 351-500, >500 cells/µL). In multivariable analysis after adjusting for age, alcohol abuse and prior CVD a positive association between immune-suppression and CVD was found. Compared to those with the most profound immune-suppression (CD4 cell count ≤200 cells/µL), those in the 200-350 cell/µL group experienced much lower rates (HR, 0.30; 0.12-0.74; p=0.009), as did the 351-500 cells/µL group (0.41; 0.20-0.85; p=0.02) and the >500 cells/µL group (0.31; 0.13-0.7; p=0.005) [209].

Carrieri et al., in the APROCO-COPILOTE also examined current CD4 count and found values <200 cells/µl were associated with increased incidence of CADE (HR 2.52; 1.15-5.48; p=0.02) in multivariable analysis when compared to CD4 cell counts ≥200 cells/µL.

Adjustment was undertaken for: sex, age, alcohol and smoking. No association was observed for CD4 cell count at enrolment (HR 0.99; 0.55–1.80; p=0.98) in the univariable analysis and the variable was not considered further [204].

Marin et al. analysed CVD deaths in CASCADE. No evidence was found for an association between time-updated CD4 cell count (HR per 100 cell increase, 0.86; 0.73-1.02) and CVD mortality incidence in the multivariable analysis. Adjustment was undertaken for: age, sex, mode, HCV-seropositivity, time-updated HIV viral load and first-line ART regimen. Nadir CD4 during the period of follow-up was examined as a categorical variable but after adjustment (as above) was not found to be associated with CVD death. When compared to those with nadir CD4 cell counts ≥350 cells/µL no association was found, after adjustment, for those with CD4 cell counts (cells/µL) 200-349 (HR, 1.43; 0.54-3.79), 50-199 (HR, 1.21, 0.41-3.59) or <50 (HR, 2.13, 0.57-7.99). A similar lack of association was observed when nadir CD4 prior to cART start was examined. Duration of immunosuppression, defined as time spent with a CD4 cell count <350 cells/µl was also not found to be associated with CVD death. When compared to those with immunosuppression of 0-1 years duration, no association was found after adjustment (as above) for those with immunosuppression for 1-3 years (HR, 1.06; 0.38-2.99), 3-6 years (HR, 1.09, 0.38-3.17) or >6 years (HR, 1.99, 0.70-5.64). Moving averages were used to reduce the impact of spurious or chance fluctuations in CD4 level in the analyses. It was not possible to adjust for a number of potential confounders not captured in CASCADE including: smoking, diabetes mellitus, dyslipidaemia and hypertension [210].

In the Data Collection on Adverse events of Anti-HIV Drugs (D:A:D) study, Sabin et al., examined the association between time-updated CD4 cell count, nadir CD4 cell count and duration of immune-suppression (time spent with CD4 cell count <200 cells/µl) and the development of various CVD outcomes. This was a large study with over 33 000 patients and well-validated end points. The analysis included 716 MI, 1056 CHD (MI, sudden cardiac death and invasive coronary procedures combined), 303 strokes and 1284 CVD (CHD and stroke combined) events. Analysis was restricted to first events. After adjustment, there was no evidence of a linear association between any CD4 related exposure and MI or CHD. An association was found with stroke and time-updated CD4 <100 cells/µl (IRR 2.26; 1.29-3.94) and 100-199 cells/µl (IRR 1.63; 1.03-2.59) when compared to those with current CD4 cell counts of 200-299 cells/µl. The authors found evidence that this association may stem, in part, from misclassification of stroke-like events. Sensitivity analysis found that inclusion of stroke-like events/rejected strokes in the adjusted analysis strengthened the association

between lower current CD4 cell count and 'stroke'. For each doubling of latest CD4 count, the IRR decreased (from 0.81; 0.74-0.89; p<0.0001) for well-validated stroke events (to 0.77; 0.71-0.84; p<0.0001) when either stroke-like events were included or rejected strokes (0.75; 0.70-0.81; p<0.0001) [211]. Also examining data from D:A:D Friis-Moller et al., in a paper primarily focusing on the association between ART class and MI incidence, also examined nadir CD4 cell count as an exposure. No evidence of an association with MI was found (IRR per 50 cell/µl increase, 0.98; 0.95-1.01) [202].

In a Danish cohort study of 2 584 virally-suppressed ART recipients, Helleberg et al., examined the association between time-updated CD4 and CVD. To reduce the impact of marked changes in CD4 count occurring by chance the study considered moving averages of the three most recent counts. A CD4 decline of ≥15% in two consecutive measurements was used as the exposure. There were 56 individuals who experienced CD4 decline. During the six months following CD4 decline, CVD rates were elevated (IRR, 11.7; 3.6-37.4). Individuals remained at elevated risk during the next six month period (IRR, 2.7; 1.0-7.5). Confidence intervals were wide as there were only three events in the first six months and four events in the next six months following CD4 decline. The study adjusted for: sex, region of origin, mode, HCV-sero-positivity baseline CD4 and time to viral suppression from HIV diagnosis. No adjustment was made for a number of potential confounders including smoking, hyperlipidaemia and diabetes. ICD codes were used to identify cases. The authors argued the results were due to reverse causality i.e. the CD4 decline was a result of elevated CVD risk and not the cause of it [208].

Analysing data from all 5 031 HIV-positive adults in Denmark, Rassmussen et al. found that a CD4 cell count ≤200 cells/µL (vs. >200) before cART start was associated with increased incidence of cerebrovascular events (CVE/stroke) (IRR 2.26; 1.05–4.86) in multivariable analysis. Adjustment included only sex, age, calendar year and country of birth. No association was found between time-updated CD4 cell count (≤200 vs >200 cells/µL) after cART initiation and CVE incidence (IRR, 1.17; 0.50–2.75) after (identical) adjustment [206].

Achhra et al. analysed data on 3 012 PLHIV from the control arms of the SILCAAT and ESPRIT trials, where all patients received cART. They examined the association between a number of CD4-related variables and CVD events including fatalities. Variables included: current, nadir and baseline CD4; CD4 slope over 3 and 7 consecutive visits; and duration of immune-suppression <200/<100/<50 cells/µL. The best predictor of MI was found to be current CD4 cell count. When both fatal and non-fatal MIs were considered together no association was seen between current CD4 cell count and MI in the adjusted analysis (HR

per $\log_2$ rise 1.05; 0.77-1.43). The analysis adjusted for: sex, age, region; ethnicity; baseline prior AIDS, baseline duration of ART, time-updated ART class and time-updated HIV viral load. When fatal events were considered alone there was an association after adjustment (HR per $\log_2$ rise, 0.63; 0.43-0.92), but this association was attenuated and no longer statistically significant when CD4 cell count was lagged by six months. End-points in the study were well-validated and there was little loss to follow-up. Data were missing on potential confounders including smoking. Median follow-up was seven years [176].

Using data from the FIRST RCT, Baker et al. examined the association between current CD4 cell count and CVD in an analysis of 1 397 individuals [110]. Just 24 events occurred. The CVD outcome included: MI, stroke, coronary revascularisation and death from chronic atherosclerotic CVD. No association was found between current CD4 cell count and CVD incidence in the univariable (HR, 0.85; 0.8-1.02) or multivariable analysis (0.91; 0.88-1.45). The multivariable model was adjusted for: latest HIV viral load, baseline age, sex, ethnicity, prior AIDS and HBV/HCV coinfection. All individuals in the study had initiated ART. This study lacked statistical power, due to the small numbers of events [110].

Lang et al., using data from the FHDH, performed a nested, matched case-control study using 289 well-validated cases of first MI. Lower nadir CD4 cell counts were present in cases when compared to controls (OR per $\log_2$ increase, 0.90; 0.83-0.97) after adjustment. The multivariable model also included: smoking, family history of CAD, hypertension, HDL cholesterol, diabetes, BMI, cocaine and/or injecting drug use, HIV viral load, exposure to PIs and CD8 cell count. No statistical difference was found in current CD4 counts (values taken 3 months before index/MI date) between cases and controls (p=0.5), but no OR was provided. [199].

Lichtenstein et al. undertook both a cohort analysis and a nested case-control (1:4 controls) study within HOPS examining factors associated with CVD [175]. The CVD outcome comprised: MI, ischaemic stroke (not haemorrhagic), CAD, angina and PAD.

In the cohort analysis baseline (defined as 01/01/2002 or first visit thereafter) characteristics were analysed in 2 005 PLHIV. This study adjusted for 10-year Framingham CVD risk, which resulted in over half of cohort participants being excluded from the study due to lack of necessary data. Baseline CD4 cell count was associated with CVD incidence after adjustment (HR per 100 cell/µL decrease, 1.08; 1.01-1.14). Adjustment accounted for Framingham 10-year risk group (four categories, low to high risk), and PWID status, alcohol use and HIV viral load, all at baseline.

In the nested case-control study nadir, baseline and current CD4 were examined. Current CD4 was defined as the closest measurement prior to the event in the preceding 12 months for cases; for controls current CD4 was defined as the value closest to the middle of the year in which the CVD event occurred (in the matched patient who had an event). One hundred and forty-eight patients had a CVD event and were each matched to four controls. In the multivariate analysis latest CD4 <350 cells/µl (OR, 3.07; 1.95-4.84; p<0.001) and 350-499 cells/µl (OR, 2.79; 1.97-4.67; p<0.001) was associated with increased odds of CVD when compared to those with a latest CD4 count >500cells/µl. As a continuous covariate latest CD4 was also associated with CVD after adjustment (OR per 100 cell/µL increase, 1.14; 1.06-1.22). For latest CD4 both as a categorical and continuous variable, adjustment was only undertaken for Framingham 10-year risk group, as no other variables (nadir and baseline CD4, baseline viral load, ethnicity, insurance status, PWID and various cumulative ART exposures) were significantly associated with the outcome. No association was found between baseline CD4 or nadir CD4 in the univariable analysis of the case-control study and so these variables were not included in multivariable analysis [175].

Data from administrative claims from 6 517 individuals with HIV at two large Boston hospitals were analysed by Triant et al. The association between both nadir and current CD4 cell counts (most recent value before censoring for all individuals) and acute MI (AMI) was explored using logistic regression. A current CD4 cell counts <200 cells/µl (vs. ≥200 cells/µL) was associated with increased odds of AMI (OR, 1.74; 1.07-2.81; p=0.02) in the adjusted analysis. Factors adjusted for in multivariable analysis comprised: age, sex, ethnicity, diabetes, lipids, HIV viral load, chronic kidney disease, smoking, time since ART start and various ART exposures. An association with AMI was also found when current CD4 was included as a continuous covariate (as was HIV viral load) in an otherwise identical multivariable model (OR per 50 cells/µL increase, 0.93; 0.89-0.97; p=0.002). When nadir CD4 replaced current CD4 in the multivariable model (and viral load peak replaced viral load) the association for nadir CD4 was of borderline significance (OR per 50 cell/µL increase, 0.95; 0.89-1.01; p=0.09) [203].

Phillips et al., in a post-hoc analysis of data from the SMART RCT, found no association between nadir CD4 at baseline (when all were ART naïve) and the development of CVD in either the DC arm (HR 1.01; 0.85-1.19) or the VS arm (HR 0.90; 0.72-1.12) of the trial. No association was found between current CD4 cell count and CVD incidence (HR per 100 cell/µL increase, 1.11; 0.99-1.25; p=0.08) when data from both arms of the trial were pooled [201].

In a nested case-control study, undertaken in the SHCS, Bucher et al. examined the association between nadir CD4 cell count and the odds of a coronary event, using the CHD definition from the D:A:D study. The study matched 98 cases to 392 controls on age, sex and smoking status [200]. Nadir CD4 cell count was found to be independently positively associated with CHD after adjustment (OR per 100 cell increase, 1.33; 1.09-1.53) i.e. those with higher nadir CD4 cell counts also had higher odds of CHD. The analysis was adjusted for: lipids, blood pressure, centripetal obesity, DM, family history, PWID, log HIV viral load and years of PI and abacavir exposure [200].

In an analysis including over 27 000 PLHIV in the US, Freiberg et al. examined the association between current CD4 cell count and acute MI in VACS-VC. Multivariable analysis adjusted for: age, sex, ethnicity, hypertension, lipids, smoking, statin use, HCV, renal disease, BMI, cocaine use and alcohol abuse. When compared to uninfected individuals, in multivariable analysis both those with CD4 cell counts <200 cells/μL (HR, 1.88; 1.46-2.40) and ≥200 cells/μL (HR, 1.43; 1.21-1.69) experienced higher rates of acute MI. Uninfected individuals were not exposed to ART and so this factor could not be adjusted for in the analysis [207].

2.4.3.1.3 The association between HIV viral load and CVD

Few studies have examined the association between HIV viral load and the development of CVD/CVE.

The association between HIV viral load at baseline (ART initiation) and CVD death was examined in the ART-CC analysis. An association was found between baseline viral load as a binary covariate and CVD death (HR for viral load >5 vs ≤5 $\log_{10}$ copies/mL, 1.54; 1.05-2.27) in multivariable analysis. This analysis adjusted for: age, sex, PWID, CD4 cell count at ART start, prior AIDS, cohort and year of ART start.

Carrieri et al., in the APROCO-COPILOTE analysis reported that having a detectable HIV viral load, as a time-updated covariate, was not associated with CADE in the univariable analysis (HR, 0.99; 0.53-1.84; p=0.98). Similarly, having a detectable HIV viral load at baseline (enrolment) was also not associated with CADE (HR, 0.99; 0.55-1.80; p=0.98). Those with detectable viral load were compared to those with undetectable viral load in both cases. These variables were not included in the multivariable analysis due to their lack of univariable association [204].

In the CASCADE analysis Marin et al. examined time-updated viral load and found a positive association with CVD death after adjustment (HR, 3.86; 1.57-9.51; p=0.003). No association

was found between viral load measures before starting cART and CVD death in an adjusted model (HR 1.68; 0.74-3.81;p=0.22) [210].

Friis Moller et al. found no association between maximum viral load and MI (IRR per log_{10} increase, 1.06; 0.95-1.18) in analysis of D:A:D data [202].

The nested case-control study undertaken by Lang et al. for the FHDH, found an association between latest (within 3 months of index/MI) viral load >50 copies/mL (vs. ≤50 copies/mL) and MI (OR, 1.51; 1.09-2.10) after adjustment for multiple risk factors (previously described). When included in adjusted analysis as a continuous variable it was also associated with the outcome (OR per log_{10} increase 1.08 (1.02-1.14)) [199].

In the analysis undertaken by Lichtenstein et al. in HOPS no association was found between baseline log viral load and CVD in univariable analysis (OR per log_{10} increase, 1.03; 0.93-1.13; p=0.62). The variable was not included in multivariable analysis due to this lack of statistical significance [175].

Triant et al. also examined HIV viral load, but data were only available for about half of all patients. In the univariate analysis a positive association was found between current VL>100 000 copies/mL and AMI (OR, 2.23; 1.37-3.65). In the multivariate model (adjustment previously described) the association was attenuated and no longer statistically significant (OR, 1.63; 0.91-2.93, P=0.10) [203].

Phillips et al. found no evidence that the increased rates of CVD in the DC arm of the SMART study were due to higher current HIV viral loads after adjustment (HR 0.83 per log increase; 0.66-1.06; p=0.1), nor was there evidence that viral load was associated with CVD events when combining data from both arms, (HR, 1.09; 0.92-1.30) [201].

Bucher et al. found that HIV viral load was independently associated with CHD in the SHCS analysis (OR per log_{10} increase, 1.44; 1.20-1.72).

2.4.4 The selection of *a priori* confounders

My subsequent analysis was restricted to MI. There are a large number of well-established risk factors for MI [212, 213]. Many of these are known to impact CVD/MI risk in those with HIV [214-216]. In addition PLHIV are exposure to ART including indinavir, abacavir and lopinavir which have been found to be associated with increased CVD risk [200, 202, 217]. Co-infections (which upregulated inflammation and immune activation) have also been found to be associated with MI risk including HCV, cytomegalovirus (CMV) and other herpes viruses [144, 147]. All these factors have the potential to act as confounders. Which of these factors are associated with my exposures of interest is not well-established. I was

missing data on a number of risk factors for MI. I included the following *a priori* potential confounders based on evidence from the literature review and on which I had data: age, sex, ethnicity, duration of indinavir, duration of lopinavir and current abacavir.

2.4.5 Recent Literature

As I went on to examine only MI as an outcome and not CVD as a whole, the update to my literature review also only includes MI. There have been a number of relevant MI publications published since 31/12/2013, which I summarise in Table 2.2b. Studies which only examined exposures at baseline [218] or all CVD outcomes combined have been excluded [219] as these were not comparable with my analysis results.

Kaiser Permanente (KP) [220] found an independent association between the severity of prior immunodeficiency (but not HIV viraemia/duration of known infection) and MI.

Those with nadir CD4 cell counts ≥500 cells/μL experienced similar rates to HIV-negative health plan members after adjustment (IRR, 0.85; 0.55-1.33; p=0.002), as did those with current CD4 ≥500 cells/μL (IRR, 1.18; 0.96-1.45; p=0.03). Those with nadir CD4 <200 cells/μL (IRR, 1.74; 1.47–2.06) and 200-499 cells/μL (IRR, 1.30; 1.07–1.58), however, experienced elevated rates after adjustment, as did those with current CD4 <200 cells/μL (IRR, 1.76; 1.31–2.37) and 200-499 cells/μL (IRR, 1.59; 1.34–1.90). CD4 cell count, nadir CD4 cell count, current HIV viral load, and duration of known HIV infection were also examined in PLHIV. Current CD4 (per 100 cell/μL increase) as a linear covariate was not found to be independently associated with MI (IRR, 1.03; 0.97-1.10; p=0.3), but nadir CD4 (per 100 cell/μL increase) was (IRR, 0.88; 0.81-0.96; p=0.006). No independent association was found for current HIV viral load (IRR, 1.03; 0.97-1.08; p=0.38) or duration of known HIV infection 5-9.9 years (IRR, 1.05; 0.77-1.45; 0.75) and ≥10 years (IRR, 0.92; 0.67-1.27; p=0.6) were compared to those infected <5 years.

NA-ACCORD examined the association between both current HIV viral load and current CD4 cell count and (well adjudicated type-1) MI (T1MI) [121]. The IRR increased with decreasing CD4 cell count as follows (vs ≥500 cells/μL): 350–499 (IRR, 1.32; 0.98-1.77); 200–349 (IRR, 1.37; 1.01-1.86); 100–199 (IRR, 1.60; 1.09-2.34) and <100 (IRR, 2.19; 1.44-3.33). Adjustment was undertaken for the following time-fixed factors: sex, ethnicity, mode of HIV infection, calendar period of enrolment and smoking (ever). Time-updated variables also included in the model were: current age, hypertension, DM, total cholesterol, LDL cholesterol, dyslipidaemia treatment and HIV viral load. After similar adjustment (including current CD4 cell count) current HIV viral load was not found to be associated with T1MI (IRR for ≥400 vs.

<400 copies/mL, 1.20; 0.92-1.56). However, in sensitivity analysis where CD4 cell count was omitted, but the model was adjusted for all other variables, current HIV viral load was associated with an increased rate of T1MI (IRR, 1.36; 1.06 to 1.75).

MI incidence in PLHIV in the US was also investigated by Salinas et al. using VACS data [221]. Baseline (180 days post cART initiation), current and cumulative measures of both HIV viral load and CD4 cell count were analysed. Adjustment was undertaken for baseline: age, DM, cholesterol, LDL, HDL, smoking, hypertension, HIV viral load (for CD4 measures) and CD4 cell count (for viral load measures). Time-updated calendar period was also adjusted for.

There was no evidence of a trend for increasing MI incidence with higher current HIV viral loads. When compared to those with viral loads ≤200 copies/mL those with current values 201-999 (HR, 1.71; 95%CI, 1.06-2.75) experienced higher rates, but those with values 1000-9999 (HR, 1.11; 95%CI, 0.64-1.93) and 10 000+ (HR, 1.30; 95%CI, 0.85-1.99) did not (final model). Those with baseline HIV viral loads ≥100 000 copies/mL did, however, experience increased incidence in the fully-adjusted model when compared to those with values <100 000 (HR, 1.41; 95%CI, 1.05-1.91). There was a trend for increasing MI incidence with increasing viraemia copy-years (VCY) after full adjustment. When compared to those with <1 000 copy-years/mL the rate increased as follows: 1000-14999 (HR, 1.61; 95%CI, 1.06-2.44), 15000-99999 (1.67; 95%CI, 1.07-2.61) and ≥100 000 (2.02; 95%CI, 1.30-3.14).

When CD4 counts were examined, the incidence of MI in those with baseline CD4 cell counts <200 vs. ≥200 cells/µL, was similar after adjustment (HR, 1.11; 95%CI, 0.82-1.49). No evidence was found that cumulative time-updated CD4 cell count was associated with MI. When compared to those with ≥2700 cell-years/µL the following HRs were reported after adjustment: 1500-2699 (1.10; 0.73-1.65), 815-1499 (1.16; 95%CI, 0.73-1.85) and <815 cell-years/µL (1.22; 95%CI, 0.73-2.03).

Table 2.2b: New literature examining MI incidence published since my literature review (01/01/2014)

			Study characteristics						Potential confounders on which data were captured*													Exposure	
Study	First author, year, [paper]	Outcome	Follow-up period	Study type	Number of PLHIV (N)	Numbers of events in PLHIV (n)	Morbidity or mortality	PYFU (median)	Ethnicity	Mode of Infection	CVD/CVE in family	Prior CVD/CVE	Smoking	BMI	cART exposure	Diabetes Mellitus	Renal failure	Hypertension	Lipids/lowering	Alcohol	Cocaine	HIV Viral Load	CD4 cell count	Duration of HIV infection
Kaiser-Permanente (USA)	Silverberg 2014 [220]	AMI	1996-2009	C	22 081	280	B/T	99 090	√	√	x	x	√	√	√	√	x	√	√	√	x	√[1]	√[1,2]	√
NA-ACCORD (North America)	Drozd 2017 [121]	MI (T1)	1995-2014	C	29 169	335	B/T	131 534	√	√	x	x	x	x	x	√	√	√	√	x	x	√[1]	√[1]	x
VACS (USA)	Salinas 2016 [221]	AMI	2006-2012	C	8 168	196	B/T	53 861	√	x	x	√	√	x	x[5]	√	√	√	√	x	x	√[1,4]	√[1,4]	x

PLHIV-People living with HIV *All adjusted for age (and sex if not restricted to one sex) [1] Current (time-updated) CD4 [2] Nadir CD4 [3] Baseline CD4 [4] Cumulative

2.4.6 Summary

Only one study examined the association between HIV infection duration (from diagnosis) and CVD (CADE) and found no association. No clear picture emerged from the literature review regarding the association between markers of immune-suppression or HIV viraemia and CVD/MI. Some studies found an association and others did not and the size of associations varied. There was much heterogeneity amongst studies with respect to exposure, outcome and adjustment which may explain the difference in findings.

2.5 Diabetes Mellitus

2.5.1 Introduction

This literature review examines the evidence for an association between duration/markers of HIV infection and DM.

2.5.2 Methods

Specific inclusion criteria, in addition to those already outlined in Section 2.1.2 are as follows:

Exposure: Unlike CVD or fractures, it is not so well-established whether DM occurs at higher rates in PLHIV compared to the general population/uninfected individuals. For this reason I have also included HIV as an exposure.

Outcomes: DM (all types) or type-2 DM. Type-1 DM alone was not a permitted outcome.

PubMed® advanced search: Search (("hiv"[All Fields] AND "diabetes mellitus"[All Fields]) OR ("type 2 diabetes"[All Fields] AND "hiv"[All Fields]) OR ("diabetes"[All Fields] AND "hiv"[All Fields]) OR ("hyperglycaemia" [All Fields] AND "hiv"[All Fields])

Hyperglycaemia was included in the PubMed search term as it improved the detection of relevant publications. Hyperglycaemia was not considered to be a relevant outcome however, unless the level of hyperglycaemia was compatible with a DM diagnosis (blood glucose ≥11.1 mmol/l or fasting blood glucose ≥7.0mmol/l).

Abstracts from the International AIDS Society (IAS) and AIDS conferences were searched from 2001-2013 using the search term “diabetes mellitus” in the conference abstracts search box.

2.5.3 Results

2.5.3.1 *Study Characteristics*

Figure 2.3 shows the study identification process. The search process identified 3 240 publications although only nine publications from seven studies were relevant to the review. Three studies were undertaken in the USA. A single study was undertaken in each of France, Denmark and Switzerland and one study was international. The follow-up period was from 1994-2010, with publication dates ranging from 2003-2012. All publications I identified were available as journal articles (as opposed to solely as abstracts). All publications reported analyses of longitudinal data. The SHCS study restricted the outcome to type-2 DM whilst the other studies included both type-1 and type-2 DM.

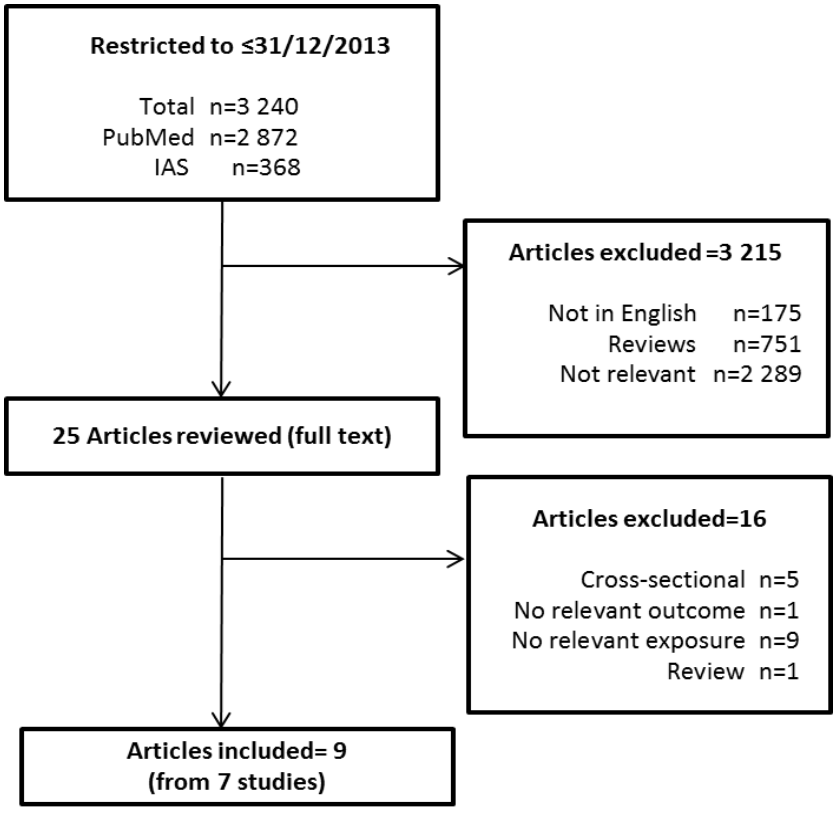


Figure 2.3: DM study selection flow diagram: Publications examining the association between markers of HIV infection and DM ≤31/12/2013

Table 2.3a summarises the characteristics of studies which explored the association between relevant markers of HIV and DM. It includes information on potential confounder data available and which covariates of interest were examined.

Table 2.3a: Characteristics of studies examining the association between markers of HIV infection and DM

		Study Characteristics				Potential confounders available to be included in final model												Covariates of interest				
Study	First author, year, reference	Follow-up	NHIV positive (N)	Events (PLHIV) (n)	PYFU (median)	Age	Sex	Mode	Ethnicity	Calendar period	BMI	cART use	Smoking	Hypertension	Central obesity	CDC stage	Triglycerides	HDL	Current CD4	Nadir CD4	Current Viral Load	Duration of infection	HIV infection
APROCO-COPILOTE (France)	**Capeau 2012 [222]**	1997-2009	1 046	111	7 846	√	√[7]	x	√	x	√	√	x	√	√	x	x	x	√[5]	x	√[5]	x	x
WIHS (US)	**Justman 2003 [223]**	1994-1998	1 435	56	3 673[6]	√	√	√	√	x	√	√	√	√	√	√	√	√	√[5]	x	√[5]	x	√
Johns Hopkins Hospital (US)	**Mehta 2003 [146]**	1996-2002	1 149	47	(0.7)	√	√	√	√	x	√	√	√	√	√	√	√	√	√[5]	x	√[5]	x	x
MACS (US)	**Brown 2005** [124]	1999-2003	229[4]	28	1 451	√	√[1]	x	x	x	√	√	x	x	x	x	x	x	x	√	x	x	x
SHCS (Switzerland)	**Ledergerber 2007** [224]	2000-2006	6 513	123	27 798	√	√	√	√	x	√[3]	x	√	√	√	√	x	x	√	√	x	x	x
	Rotger 2010 [225]	1997-2007	644	94	6 054	√	√	x	x	x	√	√	x	x	x	x	√	√	√	x	x	x	x
DHCS (Denmark)	**Rasmussen 2012** [123]	2006-2009	3 540	105	28 342	√	√	x	x	√	x	x	x	x	x	x	x	x	x	x	x	x	√
D:A:D (International)	**Petoumenos 2011** [226]	1999-2010	16 632	376	89 469	√	x	x	x	x	√	x	x	√	x[2]	x	√	x	√	x	x	x	x
	De Wit 2008 [158]	1999-2006	32 437	744	130 151	√	√	√	√	√	√	√	√	x	x	x	√	√	x	√	x	√	x

C-cohort PLHIV-People living with HIV cART-combination antiretroviral therapy HDL-High Density Lipoprotein [1]All male [2] Also adjusted for lipodystrophy [3]Dropped in multivariable model due to collinearity with central obesity [4]The analysis for HIV positive individuals on ART [5] At baseline [6]Follow-up in those with HIV [7]All female WIHS-Women's Interagency HIV study MACS-Multicentre IDS Cohort Study SHCS-Swiss HIV cohort study DHCS-Danish HIV Cohort Study D:A:D-Data Collection on Adverse Events of Anti-HIV Drugs

2.5.3.2 *The association between HIV status and DM*

There is a paucity of studies that have compared PLHIV to those without HIV infection. I only identified two relevant studies.

Using data from the Danish HIV Cohort Study, Rasmussen et al., compared DM incidence between those with and without HIV. PLHIV born in Denmark (n=4 984) were each frequency matched by age and sex to four Danish born HIV-negative controls taken from the general population and identified using a national registry. DM developed in 105 of those with HIV and 528 of those without the virus. No statistically significant association was found between HIV infection and the development of DM when the data were analysed for the whole follow-up period (1999-2010); IRR= 1.02 (95%CI, 0.83-1.26). When the data were stratified by calendar period, however, differences were noted in the association between HIV infection and DM incidence between time periods and so these were analysed separately. In the calendar period 1996-1998 a statistically significant association was found between HIV infection and the development of DM (IRR, 2.83; 1.57-5.09); when the analysis was restricted to ART-naïve individuals the association remained significant (IRR, 2.40; 1.03-5.62). In the period 1999-2009 no such association was found and those who were HIV-positive but ART-naïve were found to be at reduced risk (IRR, 0.45; 95%CI, 0.21-0.96). The advantage of this study was that the analysis used data from Denmark's comprehensive and near complete population-based cohort. This analysis therefore was representative of the whole country, had much lower loss to follow-up than most other cohort studies and controls were selected in a way likely to minimise bias [123].

An all-female US cohort study included 1 785 women, including those both HIV-positive (n=1 435) and HIV negative (n=350). The incidence of DM was only found to be elevated in PLHIV exposed to protease inhibitors when PLHIV were compared to those without the virus [223]. The analysis split PLHIV into three groups (prior PI exposure, prior ART but no prior PI and ART naïve). The incidence rate for the PI exposed group was reported as 2.8 [1.6-4.1] and for the HIV-negative group as 1.4 [0.7-2.2] cases per 100 PYFU. The other two groups (NNRTI/NRTI exposed and ART naïve) both experienced identical rates (1.2 [0.7-1.8]).

2.5.3.3 *The association between duration of HIV infection and DM*

De Wit et al. examined the association between duration of HIV infection (since first positive test) at enrolment in over 30 000 individuals enrolled in the D:A:D study [158]. In multivariable analysis the association was of borderline significance (IRR per additional

year: 0.98 [0.96 –1.00]; P=0.09). The model adjusted for: demographic factors, cohort, ART exposure, BMI, risk group, smoking and calendar year.

2.5.3.4 *The association between nadir CD4 count and DM*

Three studies examined the association between nadir CD4 cell count and DM.

Brown et al. examined the prevalence and incidence of DM in the US MACS cohort. The incidence analysis was restricted to those exposed to ART, due to low numbers of individuals who were ART-naïve. After adjustment for age, BMI and duration of cART those with a nadir CD4 cell count ≤300 cells/µl experienced significantly higher rates of DM than those with a count >300 cells/µl (RR, 1.67; 1.00-2.80). An important limitation of the study was that only 680 of the 5 622 MACS patients were included in the analysis [124].

In the SHCS, Ledergerber et al. examined factors associated with DM incidence (type 2 only). In the univariable analysis no association was found between either baseline nadir CD4 cell count <200 cells/µL (IRR 1.56; 0.86-2.82) or 200-499 cells/µL (IRR, 1.06; 0.57-1.97) when compared to those with a nadir CD4 cell count ≥500 cells/µL. Similarly, no association was found in the multivariable model after adjustment for the following baseline factors: sex, age, mode, ethnicity, CD4, CDC stage, smoking, hypertension and central (abdominal) obesity. When baseline factors (including nadir CD4) were replaced with time-updated variables (where appropriate) the results were reported as being almost identical (but these data were not provided) [224]

De Wit et al. also examined the association between nadir CD4 cell count and DM incidence in the D:A:D study [158]. The association was of borderline significance (IRR per 50 cells/µL higher: 0.98 [0.96 –1.00]; P=0.06) after adjustment (for the same factors stated above) [158].

2.5.3.5 *The association between baseline CD4/HIV viral load and DM*

Four studies examined the association between CD4 cell counts at baseline (+/- HIV viral load) and DM incidence.

Ledergerber et al., found that baseline (first visit after 1/03/2000) CD4 cell count was associated with DM development in univariable analysis. When compared to those with a CD4 cell count of ≥500 cells/µL, those with a CD4 cell count <200 cells/µL (IRR, 1.72; 1.09-2.71), but not those with a CD4 cell count 200-499 cells/µL (IRR, 0.97; 0.64-1.47), experienced higher DM rates. The statistical significance of this association was lost after adjustment: CD4 cell count <200 cells/µL (IRR, 1.48; 0.82-2.66) and 200-499 cells/µL (IRR, 0.97; 0.64-1.47) when compared to those with a CD4 cell count ≥500 cells/µL. Factors

adjusted for were the same as those previously mentioned for nadir CD4 cell count above [224].

A study by Mehta et al., at Johns Hopkins Hospital in the USA, found no statistical evidence of an association between baseline (at cART start) CD4 cell count and DM in univariable analysis amongst 1129 PLHIV [146]. However, when compared to those with CD4 cell counts >200 cells/µL, those with values 50-200 cells/µL (IRR, 1.85; 0.89-3.85) and <50 cells/µL (IRR, 1.72; 1.09-2.71) experienced elevated rates, but confidence intervals were wide, suggesting a lack of study power. Similarly, for HIV viral load, when compared to those with viral loads >10 000 copies/mL, there was little evidence that those with viral load values 400-10 000 copies/mL (IRR, 1.83; 0.85-3.95) or <400 copies/mL (IRR, 2.07; 0.72-5.98) had an increased risk of DM. Neither variable (CD4 or viral load) was included in multivariable analysis due to this lack of association.

A French study, reported by Capeau et al., examined the association between both baseline (at cART start with a PI based regimen) CD4 cell count (HR per 50/µL increase, 1.00; 0.95-1.04; p=0.87) and HIV viral load (HR comparing >500 with ≤500 copies/mL, 1.00; 0.49-2.05; p=1.00) and DM incidence in about 1 000 individuals. Due to the lack of evidence of associations in the univariable analysis both variables were excluded from the multivariable model [222].

WIHS investigators examined these associations in their all-female cohort and found that neither CD4 (HR per 10 increase in the square rooted value (HR,0.79; 0.48-1.30; p=0.4)) nor HIV viral load (HR per 1 log decrease, 1.39; 0.88-2.19; p=0.2) were independently associated with DM incidence [223]. Adjustment was undertaken for: baseline age, ethnicity, ART group, BMI and CD4 (for viral load) and viral load (for CD4). Baseline was defined as enrolment date (1994-1995).

2.5.3.6 *The association between current CD4 cell count and DM*

Two studies examined the association between current CD4 cell count and DM incidence.

In a paper by Petoumenos et al., describing the development of a model to predict the likelihood of patients developing DM in the next 6 months, the predictive effect of the inclusion of the latest CD4 cell-count was analysed. The analysis included data on 16 632 patients from the D:A:D study, 376 of whom developed DM. Compared to patients with a latest CD4 cell count <200 cells/µL, both those with counts 200-350 cells/µL (IRR, 0.52; 95%CI, 0.36-0.77; P=0.001) and those with counts >350 cells/µL (IRR, 0.51; 0.37-0.69; P=<0.001) were found to experience reduced DM rates [226].

Rotger et al., in a paper primarily focusing on the contribution of single nucleotide polymorphisms (SNPs) to DM development, examined the association between current CD4 cell count and DM in the SHCS. No statistically significant association was found between current CD4 cell count and DM development, although no numbers were given in the paper as the results were represented in a diagram; the IRR appeared to be approximately 1 with very narrow confidence intervals [177]. When baseline factors (including baseline CD4) were replaced with time-updated variables (where appropriate) in Ledergerber's multivariable analysis (also SHCS) the results were almost identical to those for baseline CD4 (this was stated in the paper but these data were not provided).

2.5.4 The Selection of *a priori* confounders

There are a number of well-established predictors of DM in the general population. These include: BMI/central obesity, age, sex, ethnicity, smoking, family history, CVD, blood pressure, steroids and polycystic ovaries [227]. In addition HCV coinfection has been associated with DM [134, 228]. ART has also been associated with DM, especially duration of PI exposure, AZT and some drugs not commonly used now (stavudine and didanosine [229-231]). My a priori confounders determined from the literature review were: age, sex, ethnicity, HCV and ART. The adjustment for cohort and calendar period was also pre-planned.

2.5.5 Recent Literature

There have been three recent relevant DM studies since my original literature review.

A US study, by Tripathi et al., examined factors associated with DM in the South Carolina (SC) Medicaid database [232]. These data were linked with a state HIV/AIDS surveillance data. PLHIV were matched 1:1 (on age, sex, cohort, ethnicity, year of cohort entry and months of enrolment) to other HIV negative SC Medicaid recipients. In multivariable analysis PLHIV who were ART-naïve were found to experience similar rates of DM to uninfected individuals (HR, 0.82; 95% CI, 0.63-1.07). Those with prior/current (>30 days) cART exposure experienced lower rates (HR, 0.55; 0.46-0.65) however. This study also examined the association between current CD4 cell count and current $\log_{10}$ HIV viral load and DM incidence, but the point estimates and 95%CIs were not reported. The publication did not reported univariable results and both variables were eliminated during backwards-stepwise selection of the multivariable model ($p>0.08$) [232].

Spagnuolo et al. examined factors associated with type-2 DM incidence after cART start at an Italian hospital [229]. Exposures included current and nadir CD4 cell count, current HIV

viral load and years since the first positive HIV test. In univariable analysis there was no evidence that current CD4 was associated with DM (HR per 100 cell/μL increase, 0.95; 95%CI, 0.89-1.10; p=0.1). An association between nadir CD4 cell count and DM was evident however (HR per 100 cell/μL increase, 0.89; 0.80-0.98; p=0.02). There was also evidence that both current log_{10} transformed viral load (HR per log_{10} increase, 1.26; 1.11-1.43; p<0.001) and current viral load as a binary variable (HR ≥50 vs <50 copies/ml, 1.52; 1.10-2.10; p=0.01) were associated with DM prior to adjustment. Duration of HIV infection (HR per 5 year increase, 0.85; 0.76-0.95; p=0.003) was also found to be associated with DM in the univariable analysis. In multivariable analysis point estimates (for variables of interest) were only provided for current CD4 count (HR per 100 cell/μL increase, 0.91; 0.84-0.99; p=0.03) and binary current viral load (HR, 2.00; 1.41-2.84, p=0.0001). It was unclear why values for nadir CD4 count and duration of known HIV infection were not reported in the multivariable results. Adjustment was undertaken for multiple factors (see Table 2.3b) including: current age, smoking (ever, ≥1 cigarette vs. none!), HCV, peak BMI (during FU) and exposure to ART (including stavudine, zidovudine and didanosine). This study found an independent association between (lower) current CD4 and (higher) current HIV viral load and elevate T2DM incidence.

A further Italian study, undertaken by De Luca et al, analysed data from the ICONA cohort [233]. They examined the association between nadir CD4 cell count and T2DM. There was evidence of an association in univariable analysis, when nadir CD4 cell counts 200-349 (IRR, 0.51; 0.30-0.86; 0.01) and counts ≥350 (IRR, 0.53; 0.36-0.77; 0.001) were compared to those with CD4 nadirs <200 cells/μL. However, after adjustment for multiple factors including sex, age, BMI, lipids, hypertension, HCV, HBV, CMV and ART the association was attenuated and no longer statistically significant. Those with nadir CD4 cell counts 200-349 200-349 (IRR, 0.66, 0.35-1.22; 0.2) and counts ≥350 (IRR, 0.67; 0.39-1.14; 0.1) were not found to be independently associated with T2DM when compared to those with nadir CD4 cell counts <200 cells/μL.

Table 2.3b: Characteristics of recent studies which examined the association between various markers of HIV infection and DM

		Study Characteristics					Potential confounders available to be included in final model													Covariates of interest				
Study	First author, year, paper	Follow-up period	Study type	HIV positive (N)	Events (n) in PLHIV	Person years of follow up (median)	Age	Sex	Mode of infection	Ethnicity	Calendar period	BMI	cART use	Smoking	Hypertension	Obesity	HCV	Triglycerides	HDL	Current CD4	Nadir CD4	Current viral load	Duration of known HIV infection	HIV infection
South Carolina Medicaid (US)	**Tripathi 2014 [232]**	1994-2011	C[1]	6 816	491	88 359	√[2]	√	x	√	√	x	√	√[2]	√	√	√	√	x	√	x	√	x	√
San Raffaele Hospital (Italy)	**Spagnuolo 2017 [229]**	1991-2014	C[1]	6 195	235	(9.8)	√	√	√	√[3]	√	√	√	√	x	x	√	√	√	√	√	√	√	x
ICONA (Italy)	**De Luca 2017 [233]**	1998-2014	C	6 505	140	38 062	√	√	√	x	√	√	√	x	√	x	√	√	x	x	√	x	x	x

[1] Cohort constructed from a hospital database [2] Time-fixed [3]Not included in the multivariable model as almost all those with DM were white

2.5.6 Summary

The three studies which compared DM incidence in individuals with and without HIV, did not find evidence that rates were elevated in those with HIV.

Few studies examined markers of HIV infection and DM incidence, but those that did found little evidence of any associations.

3 Chapter 3: Case Definitions and Data Acquisition

3.1 Survey

3.1.1 Aims

Prior to my project, CASCADE did not capture data on SNAEs. At project inception, I did not known what information collaborating cohorts were collecting on these events. One of my first tasks was to devise a survey to send to cohort investigators to gather this information. Individuals contributing data to CASCADE have a well-estimated time of seroconversion. For most contributing cohorts, these individuals come from a larger cohort population which includes sero-prevalent patients ineligible for CASCADE. This is not true for SEROCO and the UK Register which only include individuals with well-estimated seroconversion times [234].

I was interested in capturing the following information in the survey:

i. **Which SNAEs each cohort captured and approximately how many events of each type had occurred in CASCADE patients**

   It was important for me to ascertain which events were commonly captured by cohorts and how many events of each type had been recorded. I could then determine which events were likely to occur in sufficient numbers for me to undertake meaningful analysis.

   As CASCADE patients make up only a small proportion of individuals in most contributing cohorts total cohort numbers are not a good guide of how many individuals are enrolled in CASCADE and have SNAE data.

ii. **Whether standardised case definitions were used and what information was captured on each type of SNAE**

   One of my project aims was to formulate usable case-definitions for relevant SNAEs within CASCADE. I needed to ascertain what data were being captured on individual SNAEs by cohorts, whether they used standardised case definitions and how these were defined. I was also interested in coding systems cohorts used to classify SNAEs. I wished to explore the feasibility of classifying cases based on codes, for example using the Statistical Classification of Diseases and Related Health Problems (ICD). I also wanted to determine whether cohorts without case definitions were capturing data components commonly used to classify cases in cohorts with case definitions.

iii. Whether data were recorded on additional potential confounders

I wanted to assess the feasibility of gathering additional information on known risk factors for individual SNAEs not captured by CASADE at the start of my PhD. These factors included: smoking, lipids, blood pressure, BMI and family history of MI or stroke when <50 years of age.

iv. Whether cohorts already contributed to other projects which used standardised case definitions and what data were being captured for these projects

There are two collaborations (D:A:D and NA-ACCORD) which use high-quality standardised case definitions and case validation and which receive data from some CASCADE cohorts.

At the beginning of my project D:A:D [235] was already capturing information on: MI, stroke, invasive cardiovascular procedures, DM, non-AIDS cancer, end-stage liver disease (ESLD) and end-stage renal disease (ESRD). D:A:D divides cases into probable and confirmed. It uses end-point review to ensure the consistency and validity of case classification. Enrolment forms include data on prior occurrences of events and family history. Events occurring during follow-up are recorded on detailed case report forms (CRFs). Data are collected on established risk factors including smoking, lipids and blood pressure. All these features ensure that data quality and validity are high.

The North American AIDS Cohort Collaboration on Research and Design (NA-ACCORD) has extremely robust procedures and well-validated cases for: MI, non-AIDS cancer; ESRD and ESLD [11, 121, 236, 237].

v. How complete SNAE data were

I wanted to determine whether each cohort only captured data on SNAEs treated at the clinic/hospital where a patient was seen for their HIV treatment or whether additional capture of SNAE data occurred. If additional data were captured, then I wanted to find out how it occurred (self-report, linkage of medical records, registry linkage etc.).

vi. Whether data validation was undertaken and what it comprised

I wanted to know how cohort investigators ensured that their cohort data were as accurate as possible.

vii. Whether cohorts had published studies on these outcomes

I wished to learn whether cohort investigators had undertaken their own analyses in this area, whether they had published, or whether they planned to publish in the future. Some of these publications I identified during my literature review, but this aided my "grey-literature" search.

I ascertained, after discussion with my supervisors, that a number of cohorts were not capturing data on SNAEs. These cohorts were the: Greek Haemophilia cohort, Genital Shedding Study, International AIDS Vaccine Initiative (IAVI), Royal Free Hospital Haemophilia cohort, Sydney Primary HIV Infection cohort and Sydney AIDS Prospective Study.

3.1.2 Case definitions

I wanted to determine how individual SNAEs were commonly defined by clinicians and by previous observational studies and RCTs. This would give me a better understanding of how the investigators of CASCADE's contributing cohorts might be defining SNAEs and how feasible it would be for me to formulate case-definitions for SNAEs in CASCADE. I therefore undertook a number of searches during January 2013.

Firstly, I searched CASCADE cohorts' websites (for those cohorts who had one) and their publications to ascertain if they were capturing SNAEs and what standardised case definitions were used, where relevant. Secondly, I searched for publications from studies included in my literature reviews, for details of how these studies defined SNAEs. Thirdly, I undertook a PubMed search to find case definitions used in other observational studies and clinical trials; this search was undertaken on 18/01/2013. Finally, I performed a web search of relevant medical associations to determine how events were typically defined by clinicians and what information was needed for diagnosis.

Table 3.1 lists the cohorts, collaborations and clinical trial groups identified as having published data on SNAEs in patients with HIV or who stated that they were collecting SNAE data on their website.

Table 3.1: Cohorts, collaborations and clinical trials identified as collecting SNAE data

ICONA*	**SUN**	**D:A:D**
CoRIS*	**FIRST**	**MACS**
ATHENA	**ART-CC†**	**NA-ACCORD†**
EuroSIDA	**FHDH***	**HOPS**
Italian MASTER cohort	**Kaiser Permanente**	**SHCS***
INSIGHT study group	**Danish HIV cohort**	**WIHS**
VACS-VC	**CHAMPS**	

*Cohorts providing data to CASCADE †Collaboration ICONA-Italian Cohort of Antiretroviral Naïve Patients, CoRIS-Cohort of the Spanish HIV Research Network, ATHENA-AIDS Therapy Evaluation in the Netherlands, MASTER-Standardized Management of Antiviral Therapy, INSIGHT-International Network for Strategic Initiatives in Global HIV Trials, VACS-VC-The Veterans Aging Cohort-Virtual Cohort, SUN-Study to Understand the Natural History of HIV/AIDS in the Era of Effective Therapy, FIRST-Flexible Initial Retrovirus Suppressive Therapies ART-CC-Antiretroviral Therapy Cohort Collaboration, FHDH-French Hospital Database on HIV, CHAMPS-Chronic Hepatitis C Management to Improve Outcomes study D:A:D-Data collection on Adverse events of Anti-HIV Drugs, MACS-The Multicentre AIDS Cohort Study NA-ACCORD-The North American AIDS Cohort Collaboration on Research and Design, HOPS-HIV Outpatient Study, SHCS-The Swiss Hospital Cohort Study, WIHS-Women's Interagency HIV Study

Of these 20 cohorts, collaborations and trials, I found case definitions for SNAEs for:

-D:A:D (in the Manual of Operations [238])

-EuroSIDA (in the List of Definitions [239])

-SHCS (in the Definitions: non-AIDS defining events [240])

For the PubMed search I used the search terms, "diagnostic criteria" and "non-AIDS", which returned thirty-three results, only one of which was relevant. This was published by the INSIGHT Endpoint Review Committee Writing Group and described standardised case definitions for twelve Serious non-AIDS Events to be used in INSIGHT clinical trials [19]. A number of other PubMed searches were also undertaken with each combination of "diagnostic criteria" or "case definition" and the following: "myocardial infarction"; "non-AIDS malignancy"; "end stage liver disease"; "end stage renal disease"; "diabetes mellitus"; "fracture"; "pulmonary embolism"; "deep vein thrombosis"; "pancreatitis"; "congestive heart failure"; "peripheral arterial disease"; "coronary revascularization" and "stroke", but no useful publications were identified.

I searched the websites of the following medical associations and other bodies during January 2013: The American College of Cardiology [241]; American Diabetes Association [242], British Society of Gastroenterology [243], American College of Gastroenterology

[244], World Health Organisation [245], American Heart Association [246], American Stroke Association [247], European Renal Association [248], American Society of Nephrology [249], International Osteoporosis Foundation [250], European Society for Medical Oncology [251], American Society of Clinical Oncology [252] and National Institute of Health [253]. These provided me with information on common criteria used by clinicians to diagnose the following events: primary and secondary MI, ischaemic and haemorrhagic stroke, type 1 and2 diabetes mellitus, end-stage renal disease (ESRD), end-stage liver disease (ESLD), malignancies and both fragility and traumatic fractures.

During these searches I also accessed the D:A:D [235] and NA-ACCORD websites which both listed participating cohorts. The Southern Alberta Clinic was the only one of the CASCADE cohorts contributing to NA-ACCORD. ICONA, Aquitaine, the Swiss HIV Cohort Study (SHCS) and the Austrian HIV Cohort Study (through its participation in EuroSIDA) contributed to CASCADE and D:A:D.

3.1.3 Drafting the pilot survey

I drafted an electronic pilot survey in Microsoft Word®, using the survey aims and the information I had gathered on case definitions, to guide me in formulating questions.

The amount of free text in my survey was deliberately limited by the use of check boxes. Some opportunities for free text were provided, however, to permit appropriate answers if available options were not relevant to that cohort. I provided written instructions throughout the survey on which sections required completion, which could be skipped depending on responses provided and which question/section to move to next. A survey question is shown in Table 3.2a, to illustrate the survey format.

Table 3.2a: An example survey question illustrating the survey format

5. Which of the following means do you use to collect information on SNAEs for patients enrolled in CASCADE?		
Yes	No	
☐	☐	A nationwide/regional computer database is accessed from which data is extracted
☐	☐	Clinics/hospitals send datasets electronically at regular intervals
☐	☐	Clinics/hospitals send a case-report form (CRF) for each Serious Non-AIDS event
☐	☐	Clinics provide regular patient follow up questionnaires
☐	☐	Clinics provide historical data on previous SNAEs e.g. through questionnaires
☐	☐	Data are cross-checked with a national death register
☐	☐	Other – please explain the information collection process in the free text box below

The first section of the pilot survey consisted of 11 questions. These addressed various aspects of all survey aims (i-vii above). Questions included the following: were any SNAE data captured; were there case definitions for any SNAEs; did the cohort contribute to D:A:D; how were SNAE data recorded (e.g. in text-free format); was data cleaning undertaken; what, if any, coding system for events was used and had the cohort undertaken SNAE research of their own.

The second section of the pilot survey consisted of 12 sub-sections, each specific to one SNAE. The following SNAEs were included: myocardial infarction (MI); diabetes mellitus (DM); coronary revascularisation; decompensated liver disease; end-stage renal disease (ELRD); non-AIDS defining cancers; fractures; pulmonary embolisms; peripheral arterial disease (PAD); stroke; deep vein thrombosis (DVT) and acute pancreatitis.

The questions in each sub-section mostly concerned survey aims ii and v. The following information was requested (where applicable): were any data on the SNAE recorded; when did the cohort start systematically collecting data on that SNAE; how many events had been recorded; during how many years of follow-up had recorded events occurred (if known); what information on each event was captured; were case-definitions used for the event; what did case definitions comprise of and how complete were data estimated to be.

Questions regarding components of case-definitions and details of what information was captured on each SNAE, were very specific. This is illustrated in table 3.2b: a question regarding ESLD.

Table 3.2b: An example survey question regarding components of case-definitions

7. Which of the following form part of your Case definition?
☐ Evidence of cirrhosis on histological samples obtained from autopsy or biopsy
☐ Evidence of cirrhosis on MRI or CT scan
☐ Ultrasound imaging consistent with cirrhosis
☐ Hepato-renal syndrome
☐ Ascites without an alternative explanation
☐ Hepatic encephalopathy
☐ Bleeding from gastric or oesophageal varices
☐ Clinical evidence of spontaneous bacterial peritonitis
☐ Liver Transplantation
☐ Diagnostic codes (e.g. ICD-9/ICD-10)
☐ Other – please specify

The third section of the pilot survey asked about the capture of additional data on risk factors for SNAEs. Questions related to smoking, lipids, blood pressure and weight and asked whether data were collected at baseline and/or during follow-up.

The draft pilot survey was reviewed by both my PhD supervisors (KP and CS), who recommended a number of minor revisions, which were made.

3.1.4 Survey Piloting

I piloted the survey on the UK Register of HIV Seroconverters (UK Register) which was based at the MRC Clinical Trials Unit, where I was based. I sent the survey to the UK Register data manager at the end of January 2013 and they returned it within one week.

The main feedback from the UK Register data manager was that the survey was too long and had insufficient information on data cleaning. The data manager did, however, commented that the questions were straightforward and easy to understand.

On the basis of this feedback, I split the survey into two parts, with the second part only being sent out to cohorts which were found through the first part of the survey to be capturing data on SNAEs. I also tailored the second part of the survey to each cohort, so they were only sent questions about individual SNAEs they had indicated that they were capturing. The first part met (at least some components) of all (i to vii) of the survey aims. The second part concentrated on case definitions, the specific data items being recorded on each event and the number of events (i.e. it was restricted to aim ii).

In the pilot survey there was only one question regarding data cleaning and checks, "Have you performed data cleaning and checks on these data?" with a binary outcome (yes/no) (survey aim iv). This provided little information on the likely rigorousness of data cleaning and checking. I added further questions at the suggestion of the UK Register data manager to determine whether: automated data cleaning occurred during data entry, manual data cleaning was undertaken, queries were generated and sent to clinics who responded, data tracking or auditing were undertaken. I also asked whether out-of-range, missing, invalid or logically inconsistent data were routinely checked.

Other amendments were made on the basis of discussions I had with contributors and collaborators where I raised concerns I had about the pilot survey. The following revisions were also made to the pilot survey before it was sent to cohort investigators:

a) In the pilot survey questions relating to the collection of data on additional potential confounders (survey aim iii), such as smoking, only asked whether information was recorded at baseline and/or during follow-up. I realised that further questions were needed to determine what information on these potential confounders was being captured. This was necessary to allow me to assess whether combining data on these factors from different cohorts was feasible. I added questions to determine how exposure was classified. For example, with respect to smoking status, I asked whether status was recorded as binary (yes/no, never/ever), ordinal (e.g. current/ex/never) or a continuous measure (amount smoked) and what the categories comprised.
b) I added two questions to the first part of the survey, to be answered by cohort investigator s contributing data to D:A:D (survey aim iv). Firstly, I asked what percentage of CASCADE patients in their cohort were enrolled in D:A:D. This was to

give me some information on how common cases meeting D:A:D case criteria (i.e. well-validated) were likely to be in these cohorts. Secondly, a discussion with the PI of the Austrian HIV Cohort alerted me to the possibility that some cohorts contributing to D:A:D might be filling in D:A:D CRFs for all SNAEs occurring in their cohort, even if those patients were not enrolled in D:A:D. The Innsbruck centre of the Austrian HIV Cohort was doing this at the time, and there were plans for all centres in this cohort to do so in future. It was important for me to check whether this was the case for other cohorts.

c) I added a question regarding whether central adjudication was performed for those with standardized case definitions (survey aim ii) at the suggestion of one of my supervisors (CS).

d) The pilot survey included questions on decompensated liver disease, which I had assumed was a different name for end-stage liver disease. After discussing this with a clinician, I changed the name of the condition in my survey to end-stage liver disease (ESLD). Although similar, ESLD is irreversible, whilst decompensated liver disease is potentially reversible and so these two conditions differ.

The final survey, consisting of 51 questions in part 1 and 8-10 questions for each individual SNAE in part 2. Both parts of the main survey can be found in Appendix B.

3.1.5 Survey: Part One: response and findings

Part one of the survey was sent out in February 2013 to 22 cohorts of the 28 cohorts in CASCADE. The six already known not to be capturing data were excluded. The UK Register data manager completed the survey again, in-light of the new questions. These 22 cohorts are shown in Table 3.3.

Table 3.3: Cohorts who were sent Part One of the Survey on SNAE data collection

Cohort	Cohort Abbreviation
The French Hospital Database, France	FHDH
Italian Seroconversion Study, Italy	ISS
The German cohort, Germany	GER
ICONA, Italy	ICO
PRIMO, France	PRIMO
Aquitaine, France	AQU
Madrid cohort, Spain (part of GEMES)	MAD
SHCS HIV cohort, Switzerland	SHCS
Oslo and Ulleval hospital cohorts, Norway	NOR
Valencia IDU cohort, Spain (part of GEMES)	VAL
SEROCO cohort, France	SER
CoRIS, Spain	CoRIS
AHIVCOS, Austria	AHIVCOS
AMACS, Greece	AMACS
Amsterdam Cohort Study amongst homosexual men, Netherlands	NEM
Southern Alberta Clinic, Canada	SAL
Badalona IDU hospital cohort, Spain (part of GEMES)	BAD
Barcelona IDU cohort, Spain	BAR
Amsterdam Cohort Study amongst drug users, Netherlands	NEI
Lyon Primary Infection cohort, France	LYO
PHAEDRA, Sydney	PHA
UK Register of HIV Seroconverters	UKR

21 of 22 cohorts replied. Only the Italian Seroconversion Study did not respond despite five follow-up emails. This gave an overall response rate of 95% (21/22). Of the 21, 12 cohorts reported systematically collecting data on SNAEs. These 12 were subsequently sent Part 2.

A summary of survey results for the 12 cohorts systematically collecting data on SNAEs can be found in Tables 3.4-3.6.

Table 3.4 summarizes information on D:A:D enrolment, SNAE data capture and the availability of information on risk factors not captured by CASCADE at the time.

All cohorts reported collecting data on MIs, but there was variation between cohorts as to what other SNAEs they collected. Just four cohorts, for example, recorded data on coronary revascularizations. For the four cohorts contributing data to D:A:D, the proportion of individuals enrolled in D:A:D varied from 5 to 55%. D:A:D CRFs were completed for all SNAEs occurring in D:A:D enrolled patients. In addition, the SHCS was completing D:A:D CRFs for all individuals enrolled in their cohort, regardless of whether the individual was actually enrolled in D:A:D (of whom 55% were). This was the means by which they standardised data capture on SNAEs across their cohort.

All cohorts apart from The UK Register and CoRIS were collecting data on smoking, lipids, weight, height and alcohol use. The same cohorts, with the exception of FHDH, were also collecting blood pressure measurements.

Seven cohorts were found to have case definitions for (at least some) SNAEs: SAL, FHDH, AHIVCOS, AQU,, SHCS, CoRIS and ICONA. All these cohorts subsequently provided me with details of their case definitions.

Table 3.4: Summary of SNAE data and risk factors captured by the cohorts (n=12 cohorts)

	SNAEs collected														D:A:D		Risk factors collected					
Cohort/Country* (N)[1]	MI	CHF	DM	Coronary Revascularization	ESLD	ESRD	Acute Pancreatitis	nADM	Fractures	Pulmonary Embolism	PAD	Stroke	DVT	Other	Contribute to D:A:D	% patients in D:A:D	Smoking	Lipids	Weight	Blood Pressure	Height	Alcohol
SAL/Canada (226)	√	x	√	x	√	√	√	√	x	x	x	√	x	x	x	n/a	√	√	√	√	√	√
FHDH/France (10 724)	√	x	x	x	x	√	x	√	√	x	x	x	x	x	x	n/a	√	√	√	x	√	√
AHIVCOS/Austria (414)	√	√	√	√	√	√	√	√	√	√	√	√	√	x	√	5	√	√	√	√	√	√
AQU/France (1 381)	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	~55	√	√	√	√	√	√
SHCS/Switzerland (601)	√	x	√	x	√	√	√	√	√	√	x	√	√	x	√	~55[20]	√	√	√	√	√	√
COR/Spain (439)	√	√	√	x	√	√	√	√	√	√	√	√	x	√	x	n/a	x	x	x	x	x	x
ICO/Italy (1 876)	√	x	√	x	√	√	√	√	x	x	x	√	x	x	√	35	√	√	√	√	√	√
AMA/Greece (322)	√	x	√	x	√	√	x	√	x	x	x	√	x	x	x	n/a	√	√	√	√	√	√
NOR/Norway (580)	√	√	√	x	x	√	√	√	x	x	√	√	√	x	x	n/a	√	√	√	√	√	x
PRI/France (1 391)	√	√	√	√	√	√	√	√	√	√	√	√	√	x	x	n/a	√	√	√	√	√	√
SER/France (491)	√	√	√	√	√	√	√	√	√	√	√	√	√	x	x	n/a	√	√	√	√	√	√
UKR/UK (3 273)	√	x	√	x	x	x	√	√	x	x	x	√	x	√	x	n/a	x	x	x	x	x	x

*ISS, GER, PHA, LYO, GEMES and NEM/NEI did not report collecting data on SNAEs at the time of the survey MI-Myocardial Infarction CHF-Congestive heart failure DM-Diabetes mellitus ESLD-End-stage liver disease ESRD-End-stage renal disease nADM-non-AIDS defining malignancy [7]PAD-peripheral arterial disease [8]DVT-Deep vein thrombosis [1] Numbers of patients enrolled in CASCADE in 2012 [2] 55% of patients enrolled in D:A:D, but D:A:D CRFs completed for all patients

Table 3.5 examines, in more depth, what was recorded on potential confounders (all known risk factors for certain SNAEs) not routinely captured by CASCADE.

Ten of the 12 cohorts recording SNAEs also captured data on potential confounders. CoRIS and the UK Register did not collect confounder data.

All ten cohorts recorded whether or not patients were current smokers. All but one of these, FHDH, recorded smoking status at each clinic visit. Five cohorts collected data on the number of cigarettes smoked.

All ten cohorts captured total and HDL cholesterol as well as triglycerides. All cohorts collecting lipid data recorded it at regular intervals during follow-up.

The same ten cohorts captured data on weight at regular intervals during follow-up and nine of these (not AMACS) also captured height.

Blood pressure was being recorded during follow-up by all of the ten cohorts except FHDH (the largest contributor to CASCADE).

Nine of these ten cohorts recorded some information on alcohol consumption (the Norwegian cohort did not). Four of the nine cohorts recorded a previous history of alcohol misuse as well as current misuse during follow-up. Two of the nine only recorded current misuse. All nine cohorts captured the amount consumed.

Table 3.5: Summary of data captured on potential confounders (known SNAE risk factors) (n=12 cohorts)

	Smoking				Lipids				Weight/Height		Blood Pressure	Alcohol			
Cohort	**Smoking recorded**	**Current smoker**	**Status at each clinic visit**	**Amount smoked**	**Any lipids recorded**	**Total cholesterol**	**Triglycerides**	**HDL**	**Weight recorded**	**Height recorded**	**Blood pressure recorded**	**Alcohol use recorded**	**History of alcohol misuse recorded at baseline**	**Current alcohol misuse recorded**	**Amount consumed**
SAL	√	√	√	x	√	√	√	√	√	√	√	√	√	√	√
FHDH	√	√	X	x	√	√	√	√	√	√	x	√	x	x	√
AHIVCOS	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
AQU	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
SHCS	√	√	√	√	√	√	√	√	√	√	√	√	x	√	√
COR	x	-	-	-	x	x	-	-	x	x	x	x	-	-	-
ICONA	√	√	√	x	√	√	√	√	√	√	√	√	√	√	√
AMACS	√	√	√	x	√	√	√	√	√	x	√	√	x	√	√
NOR	√	√	√	x	√	√	√	√	√	√	√	x	-	-	-
PRIMO	√	√	√	√	√	√	√	√	√	√	√	√	x	x	√
SEROCO	√	√	√	√	√	√	√	√	√	√	√	√	x	x	√
UK	x	-	-	-	x	-	-	-	x	x	x	x	-	-	-

x-not recorded √-recorded

Table 3.6 outlines data sources used by cohorts to capture data and what checks were undertaken.

There was variability in the data sources utilized. Some cohorts used regular follow-up questionnaires, usually sent to cohorts annually: the UK Register, SEROCO and PRIMO used this method of data capture. Other cohorts captured data through a variety of sources including linkage to local and national registries. Cohorts enrolled in D:A:D provided detailed event forms to D:A:D on each event (for MI, stroke, invasive cardiovascular procedures, DM, non-AIDS cancer, ESLD and ESRD).

All cohorts recording any SNAEs checked for out of range values and invalid entries during data cleaning. With the exception of the SHCS, all cohorts checked missing values. Six of the 12 cohorts, including the two largest (FHDH and UK Register), checked for logically inconsistent responses. Systematic data collection of SNAEs was reported by all 12 cohorts, but for CoRIS, PRIMO and SEROCO this was restricted to some clinics only. For the Aquitaine cohort SNAE data was only collected if the event required hospitalisation for ≥48 hours. All cohorts undertook query report generation, where cohort staff contacted clinics or used other data sources to attempt to clarify data problems.

Table 3.6: Data sources used by cohorts to capture data and data cleaning checks undertaken

	Data sources										Data cleaning checks				
Cohort	EHR/EMR	Registries	Data sets	Event forms	Regular questionnaires	Death registries	PM/pathology	Lab. databases	Other	Data in coded format*	Out of range	Missing	Invalid entries	Logically inconsistent	Systematic SNAE collection for all
SAL	√	√	√	X	√	√	√	√	x	1	√	√	√	√	√
FHDH	x	x	√	X	x	x	x	x	x	2	√	√	√	√	√
AHIVCOS	√	x	√	X	x	√	√	√	x	2	√	√	√	√	√
AQU	√	x	x	√	√	√	√	√	√	2	√	√	√	x	√≠
SHCS	x	x	x	√	√	x	x	x	x	1	√	x	√	x	√
COR	x	x	x	X	√	x	x	x	x	1	√	√	√	x	√¥
ICO	x	x	√	X	√	x	x	x	x	2	√	√	√	√	√
AMA	x	x	√	X	x	x	x	√	x	2	√	√	√	x	√
NOR	√	x	x	X	√	x	x	x	x	2	√	x	√	√	√
PRI	x	x	x	X	√	x	x	x	x	1	√	√	√	x	√¥
SER	x	x	x	X	√	x	x	x	x	1	√	√	√	x	√¥
UKR	x	x	x	X	√	x	x	x	x	1	√	√	√	√	√[a]

*Coding: whether data are recorded in coded format or free text- 1- partially coded 2- all coded ¥Some clinics only ≠For those admitted to hospital EMR/HER-Electronic health record/electronic medical record PM-post-mortem

Table 3.7 shows when systematic data collection on SNAEs began for each cohort.

Table 3.7: Date when systematic prospective data collection commenced for four SNAEs (n=12 cohorts)

Cohort	Myocardial Infarction	Diabetes Mellitus	nADM	Fractures
SAL/Canada	2005	2000	2003	N/A
FHDH/France	Cohort inception (1989)	Cohort inception (1989)	Cohort inception (1989)	Cohort inception (1989)
AHIVCOS/Austria	From cohort inception	From cohort inception	From cohort inception	From cohort inception
AQU/France	1987	1987	1985	1985
SHCS/Switzerland	2000	2000	2002	2010
CoRIS/Spain	2009	2009	2009	2009
ICO/Italy	From cohort inception	From cohort inception	From cohort inception	N/A
AMACS/Greece	1/1/2006	1/1/2006	1/1/2006	N/A
NOR/Norway	Unknown[1]	Unknown[1]	Unknown[1]	Unknown[1]
PRIMO/France	From cohort inception	1999	From cohort inception	From cohort inception
SEROCO/France	From cohort inception	From cohort inception	From cohort inception	From cohort inception
UKR/UK	Unknown[2]	Unknown[2]	Unknown[2]	Unknown[2]

[1]The PI/data manager did not respond to five emails I sent requesting this information [2]There was no record of when questions about these conditions were added to the annual follow-up forms and the data manager, statistician and PI could not remember when this occurred

3.1.6 Survey: Part Two: response and findings

3.1.6.1 *Response*

Twelve cohorts were sent Part two of the survey and ten cohorts (83%) completed and returned their responses between 3/06/2013 and 25/07/2013. Responses were not initially received from two cohorts, AHIVCOS and the Oslo and Ulleval Hospital cohorts. I arranged to call the PIs of both cohorts and I gathered the relevant information over the telephone. This meant that 12/12 (100%) of cohorts reporting systematic data collection of any SNAEs completed the second part of the survey.

An estimate of SNAE numbers was provided by each cohort in Part two of the survey and these are shown in Table 3.8. Estimates suggested that there were <50 events for ESLD, ESRD, pancreatitis and stroke. MI, DM, nADM and fractures were the most frequent events.

Table 3.8: Cohort estimates of SNAE numbers
(n=12 cohorts)

Cohort	MI	DM	ESLD	ESRD	Pancreatitis	nADM	Fractures	Stroke	Total
SAL	U	U	1	1	U	6	x	0	22
FHDH	67	x	x	x	x	145	172	x	384
AHIVCOS	7	7	11	1	14	8	75	7	152
AQU	U	U	U	U	U	U	U	U	U
SHCS	15	24	2	1	x	26	10	12	90
COR	2	x	8	5	x	18	7	0	40
AMACS	13	41	x	x	x	17	x	4	75
ICONA	U	U	U	U	U	U	U	U	U
NOR	30	U	U	U	U	U	U	U	30
PRIMO	3	15	0	10	1	15	22	8	74
SEROCO	10	25	0	10	7	40	33	3	128
UKR	29	54	x	x	10	79	x	4	172
Total	176	166	22	28	32	354	319	38	1167

E-Expected U-unknown x- no data capture on this event MI-Myocardial Infarction, DM-Diabetes Mellitus, ESLD-End-stage Liver Disease, ESRD-End-stage Renal Disease, nADM-non-AIDS defining malignancy

3.1.6.2 *Details of case definitions used by cohorts*

Of the 12 cohorts capturing SNAE data, six used standardised case definitions for one or more SNAEs for at least some patients. Questions regarding case definition components were included in part two of the survey and I also requested that cohorts send me a copy of their case definitions, which they all did. I compared case definitions for SNAEs where estimated numbers of events made feasible analysis likely. Table 3.9 and 3.10 defines cases for these four SNAEs (MI, DM, non-AIDS malignancies and fractures).

Table 3.9: Case definitions for MI used by CASCADE cohorts

<table>
<tr><th>Cohort</th><th colspan="2">Myocardial Infarction</th></tr>
<tr><td>FHDH</td><td colspan="2">Based on “Case definitions for acute coronary heart disease in epidemiology and clinical research studies” [254] (reproduced below):</td></tr>
<tr><td></td><td>“Nonfatal events
A. Definite MI
1. Evolving diagnostic ECG, or
2. Diagnostic biomarkers
B. Probable MI
1. Positive ECG findings plus cardiac symptoms or signs plus missing biomarkers, or
2. Positive ECG findings plus equivocal biomarkers
C. Possible MI
1. Equivocal biomarkers plus nonspecific ECG findings, or
2. Equivocal biomarkers plus cardiac symptoms or signs, or
3. Missing biomarkers plus positive ECG”</td><td>“Fatal events (hospitalized patients)
A. Definite fatal MI
1. Death within 28 days of hospital admission in MI cases defined in I.A
2. Post-mortem findings consistent with MI within 28 days
B. Probable fatal MI
1. Death within 28 days of hospital admission in cases defined in I.B
2. Death within 6 hours of hospital admission with cardiac symptoms and/or signs. Other confirmatory data (biomarkers, ECG) are absent or not diagnostic.
C. Possible fatal coronary event
1. Death within 28 days of hospital admission in cases defined in I.C, I.F, and I.G
2. Post-mortem findings show old infarct and/or 50% atherosclerotic narrowing of coronary”</td></tr>
<tr><td>CoRIS</td><td colspan="2">Uses The American College of Cardiology/European Society of Cardiology third universal definition of acute MI [255] (includes Troponin), which is reproduced below:
“The term acute myocardial infarction should be used when there is evidence of myocardial necrosis in a clinical setting consistent with acute myocardial ischemia. Under these conditions, any one of the following criteria meets the diagnosis for myocardial infarction:
■ Detection of a rise and/or fall of cardiac biomarker values (preferably cardiac troponin) with at least one value above the 99th percentile URL and with at least one of the following: (i) symptoms of ischemia, or (ii) new or presumed new significant ST-segment–T wave (ST–T) changes or new left bundle branch block, or (iii) development of pathological Q waves in the electrocardiogram, or (iv) imaging evidence of new loss of viable myocardium or new regional wall motion abnormality, or (v) identification of an intracoronary thrombus by angiography or autopsy.
■ Cardiac death with symptoms suggestive of myocardial ischemia and presumed new ischemic electrocardiographic changes or new left bundle branch block, but death occurred before cardiac biomarkers were obtained, or before cardiac biomarker values would be increased.
■Percutaneous coronary intervention related myocardial infarction is arbitrarily defined by elevation of cardiac troponin values (>5 × 99th percentile URL) in patients with normal baseline values (≤99th percentile URL) or a rise of cardiac troponin values >20% if the baseline values are elevated and are stable or falling. In</td></tr>
</table>

Cohort	Myocardial Infarction
	addition, either (i) symptoms suggestive of myocardial ischemia, or (ii) new ischemic electrocardiographic changes, or (iii) angiographic findings consistent with a procedural complication, or (iv) imaging demonstration of new loss of viable myocardium or new regional wall motion abnormality are required. ■ **Stent thrombosis** associated with myocardial infarction when detected by coronary angiography or autopsy in the setting of myocardial ischemia and with arise and/or fall of cardiac biomarker values with at least one value above the 99th percentile URL. ■ **Coronary artery bypass grafting** related myocardial infarction is arbitrarily defined by elevation of cardiac biomarker values (>10 × 99th percentile URL) in patients with normal baseline cardiac troponin values (≤99th percentile URL). In addition, either (i) new pathological Q waves or new left bundle branch block, or (ii) angiographic documented new graft or new native coronary artery occlusion, or (iii) imaging evidence of new loss of viable myocardium or new regional wall motion abnormality.
Aquitaine	**48 hours hospitalisation with ICD-10 code** compatible with MI. D:A:D case definitions for those enrolled in addition to 48 hour criteria.
AHIVCOS and SHCS	Case definitions for all SHCS patients and AHIVCOS EuroSIDA patients from Innsbruck are those used by D:A:D [238] and are based on the MONICA study and is defined as follows: **"Acute myocardial infarction, definitive:** Definitive* electrocardiogram (ECG) or Symptoms* together with probable ECG and abnormal enzymes (or troponin)* or Typical symptoms*, abnormal enzymes* and ischaemic/non-code-able/not available* ECG or Fatal cases with naked-eye appearance of fresh MI and/or recent coronary occlusion found at necropsy **Possible Acute myocardial infarction:** Living patients with typical symptoms whose ECG and enzymes do not place them as myocardial infarction and in who there is no conclusive evidence for another diagnosis for the attack"

Case definitions for MI (Table 3.9) were consistent for four of the five cohorts using them: FHDH, CoRIS, AHIVCOS and SHCS. The American College of Cardiology (ACC) and European Society of Cardiology (ESC) standardized case definition formed the basis of the definition for CoRIS. The FHDH used the MI case definition outlined in a published paper by Luepker et al, "Case definitions for acute coronary heart disease in epidemiology and clinical research studies" which was based on the ACC/ECC definition [254]. The SHCS and AHIVCOS used the case definition for epidemiology studies used by D:A:D based on the WHO MONICA study [238] . All five included biomarkers, ECG, and pathology findings as components of diagnosis. The Universal definition used by CoRIS required typical biomarker changes for a diagnosis (in the absence of cardiac death/invasive procedures) whereas MI could be diagnosed without them in the D:A:D and FHDH definitions. Aquitaine's definition differed somewhat from the others in that it was restricted to patients hospitalised for more than 48 hours.

Table 3.10 shows case definitions for diabetes mellitus, non-AIDS defining malignancies and fractures for those cohorts using them. Four cohorts had case definitions for DM. Approximately half of Aquitaine patients were enrolled in D:A:D and so they used the same case definitions as SHCS (all patients) and AHIVCOS (patients from the Innsbruck centre only). The only case definition which was slightly different was CoRIS, which did not include HbA1C levels in the definition.

Non-AIDS defining cancer case definitions were all based on confirmed histology.

The SHCS distinguished between low (fragility) and high (traumatic) fractures. CoRIS distinguished between fragility and traumatic fractures by location of the fracture. Low impact fractures commonly occur in the hip, vertebra, forearm and humerus. The FHDH identified fractures using relevant ICD code.

Table 3.10: Case definitions for DM, non-AIDS malignancies and fractures used by CASCADE cohorts

	Diabetes Mellitus	Non-AIDS defining Malignancy	Fracture
FHDH	N/A	Confirmed histology results or imaging (liver cancer)	ICD-10 codes
CoRIS	Fasting blood glucose of ≥126 mg/dl or blood glucose of ≥200 mg/dl after an oral glucose tolerance test	Confirmed histology results and staging of tumour	Traumatic fracture recorded or non-traumatic fracture recorded (and fracture location)
SAL	No case definition	Confirmed histology	N/A (not recording fractures)
Aquitaine	24 hours hospitalisation & ICD-10 code E08-E13 & D:A:D case definitions for enrolled patients	24 hours hospitalisation & ICD-10 code C00-D49 (excluding B21) & D:A:D case definitions for enrolled patients	N/A (no work done so far on fractures)
AHIVCOS	D:A:D definition (only for EuroSIDA patients from Innsbruck) (as shown below for SHCS)	D:A:D definition (only for EuroSIDA patients from Innsbruck) (as shown below for SHCS)	No case definition
SHCS	Based on ADA criteria: Fasting plasma glucose >7.0mmol/l (126 mg/dl) The measurement of elevated plasma glucose should be repeated at least twice (different dates) without interim normal plasma glucose levels. In the absence of information on fasting glucose levels, please describe whether the diagnosis was based on: Single value of NGSP haemoglobin A1c (HbA1c) ≥6.5% (48mmol/l) Symptoms of diabetes plus random blood glucose concentration ≥11.1mmol/l (200mg/dl) Two-hour plasma glucose ≥11.1mmol/l (200mg/dl) during an oral glucose tolerance test The diagnosis has been made elsewhere and the patient has received dietary advice/started on anti-diabetic therapy.	Diagnosis of cancer (not: AIDS defining (non-Hodgkin's lymphoma, Kaposi's sarcoma), or invasive cervical); and other than basal/squamous cell skin cancers): A. In a pathology report that established the diagnosis. B. In a hospital discharge summary/consultation note from the hospital/clinic visit during which the diagnosis was established. C. In the absence of A or B: Strong suspicion of cancer supported by: (i) evidence from radiological or other imaging technique, (ii) or biochemical assay Confirmed malignancy: A or B; Probable: C	**Fracture with adequate trauma**: If fracture plausible **Low trauma fracture**: A fracture resulting from a fall from standing height or less, at walking speed or less without additional trauma/impact (i.e. a fracture whilst riding a bicycle, during sports or falling down stairs is NOT a low impact fracture)

ICD-International Statistical Classification of Diseases, ADA-American Diabetes Association, NGSP-National Glycohaemoglobin Standardization Program, D:A:D-Data Collection on Adverse Events of Anti-HIV drugs,

3.1.7 Implication of survey findings for my analysis

3.1.7.1 *Feasibility of individual SNAE analysis based on estimated event numbers*

Numbers of events recorded by the cohorts were only estimates, but it appeared that analyses of MI, DM, nADM and fractures would be feasible.

3.1.7.2 *Case definition formulation*

Case definitions for SNAEs were similar where used, but only a small number of cohorts used them. It was not possible to formulate case definitions for CASCADE, however, for two reasons. Firstly, because those cohorts without case definitions were not capturing the data components required to classify events in a similar way to cohorts with case definitions (i.e. it was not possible to restrict events to well-validated ones for these cohorts based on information these cohorts already captured). Secondly, although it was theoretically possible to restrict my analyses to events from cohorts with well-validated cases, due to the small numbers of events I could not afford to exclude events from whole cohorts as it was likely to adversely impact statistical power. I decided that for a case to be said to have occurred, all that was required was any record of an event date and event type. This would maximise the inclusion of events in my analyses. I also decided to undertake sensitivity analyses which would be restricted to cohorts with well-validated cases where feasible.

Data on SNAEs from all cohorts capturing events were therefore to be included in my analyses. I planned to examine the impact of having no case definitions, where feasible, during sensitivity analyses.

3.1.7.3 *Capture of data on risk factors (potential confounders) not captured by CASCADE*

It appeared from the survey that capturing data on additional potential confounders might be feasible, as most of the 12 cohorts were capturing most of the confounders in a time-updated manner. These comprised: smoking, lipids, weight, height, blood pressure and alcohol use.

3.1.8 Consultation with collaborating cohorts and data capture

3.1.8.1 *The first SNAE TC*

Once I had analysed the information from the surveys, I arranged a meeting with collaborating cohorts to discuss the feasibility of capturing data on SNAEs. I contacted the PIs of all cohorts who had indicated they had data on SNAEs and invited them to a teleconference (TC). The meeting took place on 29th October 2013.

At this meeting, which eight cohort PIs were able to attend, it was decided that the collection of data on SNAEs was feasible and would be a one-off event, but might be incorporated into the annual CASCADE data merger in subsequent years.

It was agreed that data on the following SNAEs would be included: myocardial Infarction, diabetes mellitus, end-stage liver disease, end-stage renal disease, pancreatitis, non-AIDS defining malignancies, fractures and stroke. Only the event date and the type of event would be needed for an event to be included in any analysis. This was due to the small number of events and the lack of recording of necessary data components.

Concern about the lack of robust case definitions, due to the potential for misclassification was raised. Sensitivity analysis to assess the impact of case definition quality on any findings was discussed as a possible solution to this problem. A number of cohort investigators were concerned that under-reporting of events was a potential problem and the investigators of six cohorts (AHIVCOS, SAL, FHDH, UKR, PRIMO and SEROCO) noted that they had found this to be the case in their cohorts.

3.1.8.2 *Determining a suitable format for data capture*

CASCADE uses the HIV Cohorts Data Exchange Protocol (HICDEP) format for data transfer which takes the form of a series of tables [256]. At the time, HICDEP had not developed any tables for capturing data on SNAEs. It was important that the table was as simple and short as possible, to reduce the additional amount of work for data managers.

I compiled a draft table to send to collaborating cohorts using the HICDEP structure. After discussions with one of my supervisors (CS), I used the SNAE categories used by EuroSIDA and the D:A:D codes for non-AIDS defining malignancies. The table included: patient identification, patient birth date, SNAE code, SNAE data and non-AIDS malignancy code where appropriate. A copy of the table can be found in Appendix C. This table was named tbl_SNAE.

3.1.8.3 *The second SNAE TC*

The second TC was held on the 4th December 2013. The HICDEP table I had created to capture data on SNAEs was discussed and found to be acceptable to those attending. I subsequently checked with PIs unable to attend the TC and the format was agreeable to them too. Authorship rules of any publications arising from the study were also discussed CS commented that previous research in D:A:D/EuroSIDA had found that associations between exposures and malignancies varied by malignancy type and so composite SNAE end-points might not be valid. It was agreed that pooling data on different SNAEs and analysing them together was biologically implausible and so they would be analysed separately.

It was also decided that it was not feasible to capture data on additional potential confounders such as smoking and BMI. This was because it would require the data managers to provide these data for all individuals enrolled in CASCADE, whereas capturing additional data on SNAEs would only involve sending additional information (i.e. providing tbl_SNAE data) on patients with an event. It was felt necessary to reduce the work of data managers. The SHCS subsequently decided not to participate in the study due to the decision not to capture data on these potential confounders.

3.1.8.4 *SNAE Data Merger methods*

The table (Tbl_SNAE) was sent to eleven cohorts on 12/11/2013. Cohorts sent the table comprised: The UK Register, FHDH, Aquitaine, AHIVCOS, ICONA, AMACS, SAL, PRIMO, SEROCO, The Oslo and Ulleval Hospital Cohort and CoRIS.

Aquitaine cohort agreed to participate but never provided data as they had not mapped codes for each SNAE. All other cohorts provided data.

3.1.8.5 *Receipt of SNAE data and data cleaning*

I received Tbl_SNAE data from cohorts between 29/11/2013 and 01/04/2014. The usual methods of data transfer used by cohorts to transfer data to CASCADE were used in-line with country of origin specific and UK data protection law.

Tbl_SNAE was only a single table, numbers of patients with events were small and data managers generally filled it in very accurately, so the table required only limited data cleaning. SNAE data for each cohort was initially cleaned separately. After initial visualisation of all data, the data underwent a series of checks including the identification of: incorrectly formatted data files (e.g. a SNAE was coded as DAI instead of DIA), missing data (e.g. no event date), duplicate data (e.g. the same event occurring for the same

patient on the same day recorded twice), impossible values (e.g. SNAE event occurring after recorded death date), unlikely/out of range values (e.g. birth date before 1925), patients whose unique identifier could not be matched with existing CASCADE data. I sent all errors and inconsistencies as queries to data managers and data were amended accordingly.

I then merged the combined SNAE dataset using unique identifiers with the existing CASCADE 2013 dataset which had previously been cleaned by the CASCADE data manager/statistician. I used birth dates to double-check that the data being merged (from Tbl_SNAE and the main dataset baseline information Tbl_BAS) were from the same patient.

3.1.8.6 *Comparison of expected and received SNAE numbers*

Table 3.11 shows the approximate number of events reported from the second survey compared to the actual number of events received from the data merger. There was an over-estimation of the numbers of events in the second survey for the following conditions: MI, DM, ESLD and ESRD. The second survey under-estimated the numbers for: pancreatitis, nADM, fractures and stroke.

3.1.8.7 *Number of events after data cleaning*

I received data on 448 fractures and 147 MI and 147 DM (Table 3.11). Repeat events and those with missing event dates were removed. This left 300 fractures, 121 MI and 109 DM. Although information on 416 non-AIDS defining malignancies were received, it was not biologically plausible to analyse them together (see Section 2.3.3) and, due to the diversity in types of malignancy, there were insufficient numbers to analyse them individually. My analyses were therefore restricted to fractures, MI and DM, and hence the restriction. Table 3.11 included repeat events. Subsequently only the first event of each type was used in my analyses. Some cohorts inadvertently gave numbers for their entire cohort in their estimate and did not restrict the estimate to their CASCADE patients. Events which were missing an event date were also subsequently removed prior to analysis.

Table 3.11: A comparison of the number of SNAEs expected (E) based on the survey results and the number actually received (R)

Cohort	MI		DM		ESLD		ESRD		Pancreatitis		nADM		Fractures		Stroke		Total	
	E	R	E	R	E	R	E	R	E	R	E	R	E	R	E	R	E	R
SAL	U	2	U	12	1	0	1	0	U	3	6	8	x	x	0	3	22	28
FHDH	67	41	x	x	x	x	x	x	x	x	145	210	172	279	x	x	384	530
AHIVCOS	7	9	7	19	11	2	1	2	14	15	8	30	75	105	7	12	152	194
SHCS	15	n/a	24	n/a	2	n/a	1	n/a	x	x	26	n/a	10	n/a	12	n/a	90	n/a
COR	2	0	U	1	8	1	5	0	x	x	18	3	7	0	0	0	40	5
AMACS	13	0	41	7	U	5	x	3	x	x	17	2	x	x	4	6	75	23
ICONA	U	11	U	32	U	5	U	0	U	0	U	61	U	0	U	3	U	112
NOR	30	41	U	11	U	0	U	1	U	7	U	17	U	10	U	13	30	100
PRIMO	3	1	15	14	0	0	10	11	1	4	15	12	22	25	8	6	74	73
SEROCO	10	15	25	15	0	0	10	10	7	7	40	32	33	29	3	4	128	112
UKR	29	27	54	36	x	x	x	x	10	9	79	41	x	x	U	4	172	117
Total	176	147	166	147	22	13	28	27	32	45	354	416	319	448	34	51	1167	1294

E-Expected R-Received U-unknown x-not captured by that cohort MI-Myocardial Infarction, DM-Diabetes Mellitus, ESLD-End-Stage Liver Disease, ESRD-End-Stage Renal Disease, nADM-non-AIDS Defining Malignancy

3.1.9 Main Survey findings: Summary

Twelve cohorts were systematically collecting data on SNAEs and were interested in participating in the study.

Ten of the twelve were gathering data on at least some potential confounders not captured in CASCADE (smoking, alcohol, lipids, BMI and blood pressure). What information was being recorded and how long cohorts had been capturing these data was highly variable. After discussions with collaborating PIs it was decided that I could not collect data on additional potential confounders as it was too much work for cohort data managers to provide it. Information on all individuals enrolled in CASCADE would be needed, in contrast to my data capture of outcomes which only required additional information on cases.

Few cohorts used well-validated case definitions. Most that did were doing so through their collaborations with D:A:D, NA-ACCORD or EuroSIDA. For MI (where case definitions are potentially very important) I subsequently undertook a sensitivity analysis to explore the effect of case definition quality. At the time of my survey most cohorts without robust definitions did not routinely capture the additional information needed to classify cases rigorously.

My survey results indicated that there were sufficient numbers of events to examine MI, fractures, DM and nADM (if combined).

4 Chapter 4: Methods

4.1 Introduction

This chapter outlines the methods I used to examine the association between various markers of HIV infection and three different SNAEs (fractures, MI and DM).

The primary aim of each analysis was to explore the association between duration of HIV infection and event rates after controlling for other factors. The secondary aim was to examine the association between current HIV viral load and various CD4 metrics and event rates.

Cohort studies are prone to certain biases. These methods outline how I attempted to minimise bias by attending to the following issues: confounding, over-fitting, collinearity, outliers, interactions/effect modification, missing data and non-linearity.

4.2 The CASCADE cohort collaboration

4.2.1 Background

CASCADE was established in 1997 and initially comprised mostly European cohorts but also a small number from Australia. In more recent years it has grown to include cohorts from North America and Africa [257]. CASCADE's primary aim is to address important HIV-related research questions which require a large sample size and an accurate estimate of likely HIV seroconversion date. I used data from the CASCADE cohort collaboration for all analyses in this PhD.

4.2.2 CASCADE inclusion criteria

For an individual to be eligible for inclusion in CASCADE a reasonably accurate estimate of when HIV seroconversion took place is needed. These estimates are based on available laboratory data. There are three criteria by which individuals can be included in CASCADE (CASCADE File Specification version 6, 2012). Firstly, if there is laboratory evidence of current HIV seroconversion i.e. real-time PCR positivity or an incomplete Western blot. In this instance the date of the laboratory test is taken as the date of HIV seroconversion. Secondly, if an individual has a negative HIV test and this is followed by a positive test within 3 years. The seroconversion date is then taken as the midpoint between the two tests (or date of seroconversion illness if documented). Thirdly, individuals with haemophilia are eligible for inclusion based on the date they are likely to have received contaminated blood products (based on a probability distribution).

Individuals are ineligible for inclusion in CASCADE if they: contracted HIV through vertical transmission, were ≤15 years old at HIV seroconversion, became infected through blood transfusions (unless a haemophiliac), acquired HIV through hospital procedures or lacked clinic documentation of the date of the HIV negative test (i.e. was self-reported).

4.2.3 Research within CASCADE

CASCADE research to date has included examining changes in life expectancy and causes of morbidity and mortality over time. This has included studying the factors associated with changing morbidity and mortality, including the impact of ART [210, 258, 259]. CASCADE has also examined: how HIV virulence and transmissibility has changed over time [260], how an individual's genotype, or HIV virus characteristics (including sub-type or acquired resistance) affects disease progression or response to ART [261-263], how HCV co-infection incidence has changed with time and its effect on mortality [150, 264] and how HIV infection progression differs between high, middle and low income countries [265, 266].

The rapidly evolving HIV epidemic in Eastern and Central Europe has been a recent focus of study, with particular emphasis on estimating HIV incidence in Poland, Estonia and the Ukraine [267, 268]. Recent CASCADE research has also evaluated definitions for and characteristics of elite controllers (individuals who maintain HIV viral control for long periods post-seroconversion without ART) [269], long term non-progressors (individuals who maintained a high CD4 and have no AIDS events long term without ART) [270] and rapid progressors (those with rapid post-seroconversion CD4 decline) [271, 272]. In the last few years projects have also included: assessing how rates of transmitted drug resistance have changed over time [273], determining what factors predict CD4 cell recovery in individuals starting ART [274] and using phylogenetic trees to examine transmission between risk groups in Eastern Europe [275].

CASCADE is part of EuroCoord, a network of excellence made up of a number of the largest HIV cohorts and cohort collaborations in Europe. The other members of the network are The Collaboration of Observational HIV Epidemiological Research Europe (COHERE), EuroSIDA and Paediatric European Network for Treatment of AIDS (PENTA) [276].

4.2.4 CASCADE data

Up until the end of 2015, when European funding of EuroCoord through the Seventh Framework Programme ended, the CASCADE data merger usually occurred annually. Participating cohorts sent individual patient data, which were cleaned and merged to

produce the annual dataset. Table 4.1 shows the 28 cohorts who have contributed data to CASCADE.

Table 4.1: Cohorts contributing data to CASCADE

Abbreviation	Cohort and Country
AMACS	AMACS, Greece
AQU	Aquitaine cohort, France
AHIVCOS	Austrian HIV cohort study, Austria
BAD	Badalona IDU hospital cohort, Spain
BAR	Barcelona IDU cohort, Spain
CoRIS	CoRIS, Spain
FHDH	French Hospital Database, France
GER	German cohort, Germany
GRE	Greek haemophilia cohort, Greece
GSS	Genital Shedding Study, Uganda and Zimbabwe
IAV	IAVI, Kenya, South Africa, Uganda, Rwanda, Zambia
ICONA	ICONA cohort, Italy
ISS	Italian Seroconversion Study, Italy
LYO	Lyon Primary Infection cohort, France
MAD	Madrid cohort, Spain
NEI	Amsterdam Cohort Study among drug users, Netherlands
NEM	Amsterdam Cohort Study among homosexual men, Netherlands
NOR	Oslo and Ulleval hospital cohorts, Norway
PHA	PHAEDRA, Sydney, Australia
PRIMO	PRIMO Cohort, France
RYF	Royal Free haemophilia cohort, United Kingdom
SAL	Southern Alberta clinic, Canada
SEROCO	SEROCO cohort, France
SHCS	Swiss HIV cohort, Switzerland
SYO	Sydney Primary HIV Infection cohort, Australia
SYP	Sydney AIDS Prospective Study, Australia
UKR	UK Register of HIV Seroconverters, United Kingdom
VAL	Valencia IDU cohort, Spain

CASCADE routinely captures data on: demographic factors, CD4 cell counts, HIV viral load, ART and other medication, viral co-infection with HBV and HCV, causes of death, AIDS events and resistance.

4.2.5 Cohorts

Cohorts which systematically captured data on events included in these analyses are shown in Table 4.2.

Table 4.2 Cohorts which contributed data to SNAE analyses

	Myocardial Infarction (MI)	Fractures	Diabetes Mellitus (DM)
FHDH	√	√	x
AMACS	X	x	√
AHIVCOS	√	√	√
PRIMO	√	√	√
SEROCO	√	√	√
NOR	√	√	√
ICONA	√	x	√
SAL	√	x	√
CoRIS	√	x	√
UKR	√	x	√
Total	**9**	**5**	**9**

4.3 Inclusion criteria and follow-up time calculation

For each event, individuals were included in the analysis if they were enrolled in CASCADE and:

- were ≥15 years of age at estimated HIV seroconversion date with known birth date
- had at least one CD4 cell count and HIV viral load measurement during follow-up
- were from a cohort (or subset) systematically collecting data on the event
- were followed-up after the date their cohort started systematic event data capture

Contributing cohorts have historically sent annual data updates to CASCADE; the 2013 CASCADE dataset was used for all my analyses. The dataset was merged with the additional data captured on SNAEs in 2013 (see section 3.1.8 of Chapter 3 for details).

For each event a flow-diagram was created outlining reasons for exclusion from the analysis and the number of individuals and events excluded. All eligible patients were followed-up from cohort enrolment or from when the cohort started systematically collecting data on the event, whichever occurred later.

The earliest date of entry into follow-up for each event type by cohort is shown in Table 4.3. Some cohorts undertook retrospective systematic data capture prior to the official date of cohort inception and these dates are used where appropriate. The median date of

entry into follow-up [IQR] for each event type was as follows: fractures 08/07/2004 [31/03/1998-09/06/2008]; MI, 16/12/2003 [16/06/1998-12/06/2008]; and DM, 18/08/2003 [16/06/2000-27/04/2009].The analysis examined time to first known event for each event type.

Table 4.3: First date of entry into follow-up by event type and cohort

Cohort	MI	Fractures	DM
FHDH#	06/01/1992	06/01/1992	-
AMACS	-	-	02/09/1980
AHIVCOS~	26/06/1985	26/06/1985	26/06/1985
PRIMO#	06/06/1996	06/06/1996	06/06/1996
SEROCO#	12/01/1988	12/01/1988	12/01/1988
NOR*	21/08/1987	01/04/2010	02/01/1993
ICONA	13/08/1996	-	13/08/1996
SAL~	01/01/2005	-	01/01/2000
CoRIS#	15/01/2004	-	15/01/2004
UKR*	16/06/1998	-	16/06/2000
Overall	26/06/1985	26/06/1985	02/09/1980

*For the Oslo and Ulleval Hospital cohort and for the UK Register there was no information on the date that systematic data collection of individual SNAEs started; for these patients the first possible date of entry into follow-up for each analysis was imputed as the day after the first event of that type was recorded for that cohort. #Cohorts who had captured event data since cohort inception ~ The Southern Alberta Clinic had known dates at which systematic data collection commenced, but these differed by event type

Follow-up ended at the data submission date, six months after the patient's last clinic visit, death or first event, whichever happened first. Although some loss to follow-up dates and information on administrative censoring were available in the CASCADE dataset, these were not used to right-censor follow-up as the information was incomplete.

Each patient's follow-up was split at the beginning of every month post HIV seroconversion (to create an accurate measure of time since seroconversion) and then again at the date of each new laboratory measurement (CD4 cell count or HIV viral load) or change in ART/HCV/AIDS status. The status of each patient with respect to all factors was updated at the start of each period (with the most recent measure available). By definition, each period was no longer than one month. Splitting the data in this way enabled the value of factors to change with time.

4.4 Variables and statistical methods

4.4.1 Confounding and covariate variables

Table 4.4 shows the variables considered for inclusion in the analyses. The table highlights whether a variable was considered to be a covariate of interest or a potential confounder and whether it was time-updated or time-fixed. Further information on variable definitions follows in the text below the table.

Table 4.4: All variables considered for inclusion in the SNAE analyses

	Covariate of interest (I)/ Potential confounder (C)	Time-updated (U)/ time-fixed variable (F)
Age at seroconversion	C	F
Sex	C	F
Ethnicity/region of origin*	C	F
Mode of infection	C	F
Cohort name	C	F
Current age	C	U
Measures of ART exposure	C	U
Calendar period	C	U
HCV-seropositivity	C	U
Current CD4 cell count	I	U
Nadir CD4 cell count	I	U
Duration of immunosuppression (CD4 count ≤200/100/50 cells/µL)	I	U
Current HIV viral load	I	U
Prior AIDS	I	U
Duration of HIV infection	I	U

Demographic factors are highlighted in grey * Region of origin was used to determine likely ethnicity where ethnicity was missing. U-time-updated F-time-fixed I-covariate of interest C-potential confounder

Information on ethnicity was missing for a large number of individuals, mostly because collection of these data is not permitted by national law in France. I decided to use region of origin to predict probable ethnicity where ethnicity data were missing, applying the following rules:

1. If a patient's region of origin was recorded as Africa or the Caribbean they were classified as of black ethnicity.
2. If their region of origin was Europe/North America or Australasia they were classified as of white ethnicity.
3. Those from other regions were classified as of "other" ethnicity.

To assess the validity of these assumptions, agreement between ethnicity and assumed ethnicity based on region of origin were compared for individuals on whom information on both variables were available (see Table 5.2 in Chapter 5).

HIV viral load measurements were excluded from each analysis if they had occurred within three months of seroconversion as viral load is known to fluctuate widely during this time [277]. Undetectable viral load measurements were assigned half the lower limit of the assay. Where the lower limit of the viral load assay was missing, 400 copies/mL was used as the lower limit, and 200 copies/mL was assigned to undetectable values.

The distribution of all continuous variables was investigated and values truncated at the 1st and 99th percentiles. Duration of HIV infection was calculated from estimated seroconversion date (see Section 4.2.2 for details of how this was estimated). Three measures of duration of immune-suppression were calculated at the beginning of each time period, which were the total amount of time that an individual had spent with a CD4 cell count ≤200/100/50 cells/µL respectively. Nadir CD4 cell count was calculated as the lowest value recorded up to the beginning of each period.

A number of assumptions were made about HCV status. The formulation of these assumptions was influenced by the apparent sparsity of testing for most patients and incomplete data on treatment:

- once an individual had tested positive with respect to HCV (whether viral load, antigen or antibody) they were classified as HCV-seropositive from that point on
- individuals with no HCV test data were considered to be HCV seronegative

ART related variables included in each SNAE analysis (and whether they were treated as a current or cumulative measure) were specific to each event. These variables comprised potentially important exposures identified from previous studies (see relevant literature reviews in Chapter 2 for details). A time-updated categorical variable summarising current ART was generated for the fractures analysis with the following five groups: off ART, tenofovir disoproxil fumarate (TDF) but no protease inhibitor (PI), PI but no TDF, PI and TDF; and all other ART (no PI or TDF). For the MI analysis current abacavir (ABC) and duration of exposure to both indinavir (IDV) and lopinavir (as two separate variables) were included. For the DM analysis a number of ART related variables were considered: ART, nucleoside reverse-transcriptase inhibitors (NRTI), non-nucleoside reverse-transcriptase inhibitor (NNRTI) and PI (both as current and cumulative exposure); in addition, zidovudine (AZT), stavudine (d4T) and didanosine (ddI) were examined both as current and cumulative

exposures. Cumulative exposure, for all ART variables, was defined at the beginning of each time period as the cumulative duration of exposure since enrolment.

4.4.2 Main Statistical Methods

4.4.2.1 *Rates and univariable analysis*

First event rates (per 1000 person-years) were calculated as the number of first events divided by the person years at risk multiplied by 1000 for all patient follow-up and then for each level of the variables listed in Table 4.4. Two-sided 95% confidence intervals were calculated using the Normal approximation. A Poisson regression model examined univariable associations between variables and the event of interest. Results are presented as incidence rate ratios (IRRs) comparing the incidence rate for each level of the variable to the incidence in the baseline category.

4.4.2.2 *Multivariable analysis*

A multivariable Poisson regression model was built using a series of manual steps, to explore the association between duration of and markers of HIV infection (covariates of interest) and each event:

i. Confounding and over-fitting: Initially a basic multivariable model (model A0) was created containing potential confounders, including demographic factors (Table 4.4). Event-specific pre-selected potential confounders (*a priori*) identified after reviewing the literature were included where available regardless of their statistical significance. I had planned to also include any additional variables listed as potential confounders in Table 4.4 which were found to be associated with the event during subsequent backwards-stepwise selection ($p<0.05$), but all had $p\geq0.05$ and so were not retained. Each model A0 was simplified by collapsing categorical variables (where IRRs for categories were similar) and comparing Akaike's Information Criterion (AIC) scores for models including the collapsed and un-collapsed categories to ensure this was appropriate. The aim of simplifying Model A0 was to keep the number of parameters as low as possible to aid in the prevention of over-fitting (see Section 4.4.2.4). Over-fitting occurs when a model fits the data too closely and, as a result, may not fit or predict additional or future data very well.
ii. Collinearity: This was assessed between potentially related continuous variables by examining scatterplots and calculating Pearson's correlation coefficients for each pairwise association to ensure it was appropriate to include them in the model

simultaneously. The association between categorical variables was then assessed using Fisher's exact test. If variables were strongly collinear or strongly associated then only one of the variables was included in the model; the decision as to which variable to retain was based on comparing the strength and size of the associations and on their biological plausibility as predictors.

iii. Interactions: Potential interactions between each confounder in model A0 and current age, sex and ethnicity/region were assessed and any interaction with a p value <0.1 was investigated further. P<0.1 was used due to the insensitivity of the test for interaction. If stratification by the effect modifier produced IRRs where the larger one was ≥20% bigger than the smaller one, then an interaction term was included in the model. If stratified IRRs were both larger or both smaller than the combined value this was taken to indicate simultaneous confounding and effect modification. At this point model A was defined and fixed.

iv. Covariates: All covariates of interest (see Table 4.4) were added one at a time to model A. Each of these models was assigned a letter as follows:

B-duration of HIV infection (years)

C-Current CD4 count (per 100 cells/μL increase)

D- HIV viral load (per $\log_{10}$ copies/mL increase)

E-Nadir CD4 cell count (per 100 cells/μL increase)

F- Duration of immune suppression ≤200 cells/μL (per additional year)

G- Duration of immune suppression ≤100 cells/μL (per additional year)

H- Duration of immune suppression ≤50 cells/μL (per additional year)

I-Prior AIDS

v. Those found to be associated with the outcome at p<0.1 were retained with the exception of duration of HIV infection which was retained regardless of the strength of its association.

vi. All retained covariates of interest were added to model A. Covariates of interest underwent backwards stepwise selection (p<0.1) to determine their suitability for inclusion in the final model, except for duration of HIV infection which was retained in this model regardless of its statistical significance (as were all model A variables).

vii. Checking of final model: previously excluded variables were added one at a time to the final model. None were found to be associated with the outcome for any of the events (at p<0.05 for potential confounders or p<0.1 for covariates of interest).

viii. Additional interactions: duration of infection and included covariates of interest were then assessed for interaction with current age, sex and ethnicity/region (using the same approach as in step C).

ix. The final model, generated from all these steps, was labelled model J.

4.4.2.3 *Collinearity of duration of HIV infection and measures of age*

Issues of collinearity affected multivariable model building. Current age for any individual is the sum of age at HIV seroconversion and duration of HIV infection, therefore, it was only possible to include any two of these three variables in the same model. I aimed to evaluate the impact of duration of HIV infection on the risk of SNAEs, so it was essential to include this variable in the final models. Age at seroconversion and current age were strongly positively correlated but both were less strongly correlated with duration of infection (see Chapter 6, text below Table 6.4.2 for details). In multivariable models I adjusted for current age in preference to age at seroconversion for two reasons: firstly, increasing age is a well-recognised predictor of each event, secondly current age in the univariable analyses was a stronger predictor of the risk of each event than age at seroconversion.

4.4.2.4 *The prevention and assessment of over-fitting*

The risk of overfitting was low as models had more than 10 events per parameter (this 1:10 rule has been shown to be somewhat conservative [278]). The final model for each event was built using a series of steps which aimed to balance the need for excluding unnecessary variables (thus reducing the risk of over-fitting) whilst adjusting for all potentially important ones (accounting for confounding). Final models were assessed statistically for over-fitting (using the Overfit command in STATA).

4.4.2.5 *Missing data*

Multiple imputation of missing data was not undertaken. For time-varying covariates it was not possible to distinguish if a value was missing or had not been taken, but the last observation was carried forward unless otherwise stated.

4.4.2.6 *Treatment of continuous variables*

All continuous variables were initially assessed as categorical. Subsequently they were treated as continuous and in the main analysis assumed to be linear with the exception of HIV viral load which was $\log_{10}$ transformed. Deviations from linearity for all variables were explored during sensitivity analysis using multivariable fractional polynomials (See Section 4.5).

4.4.2.7 *Poisson regression model: assessing the validity of assumptions*

The Poisson model makes a number of assumptions [279]: events are rare, independent, occur at a constant frequency and the mean and variance are equal. My outcome was rare and independent and likely to have occurred at constant frequency. If there is over-dispersion (variance greater than mean) then a negative binomial model better describes the data. If there are excessive zeros (fewer events than predicted by the model) then a zero-inflated model is needed. I compared zero-inflated, negative binomial and Poisson models to see which best fitted my data.

4.5 Sensitivity analyses

I performed a number of sensitivity analyses for each event to test the robustness of assumptions.

4.5.1 Sensitivity model 1a/1b: assessing the effect of carrying forward CD4 cell counts

In sensitivity model 1a individuals who had one or more gaps of ≥1 year between CD4 measures were censored at the first gap of ≥1 year. A diagrammatic representation of this sensitivity analysis is shown in Figure 4.1. Nadir CD4 cell count and all measures of duration of immune-suppression were recalculated using the amended CD4 data. Model A from the main analysis was refitted (i.e. the model was not rebuilt) to the censored dataset and each CD4-related covariate added to model A in turn (models C, E, F, G, H). Results were compared to models C-H of the main analysis.

Sensitivity model 1b was similar to 1a, but individuals were allowed to re-enter follow-up if a subsequent CD4 measurement occurred. They then remained in follow-up until the end of follow-up or until there was a second gap of ≥1 year between CD4 measurements when censoring occurred again. It was possible for patients to contribute multiple distinct episodes of follow-up if multiple gaps ≥1 year occurred. This approach was taken to increase follow-up (and therefore power) over Model 1a. Model A was refitted to the amended data and covariate results compared.

4.5.2 Sensitivity model 2a/2b: assessing the effect of carrying forward HIV viral load measurements

Sensitivity model 2a /2b were similar to 1a/1b but explored the impact of censoring the follow-up of individuals who had one or more gaps of ≥1 year between HIV viral load measurements (Figure 4.1).

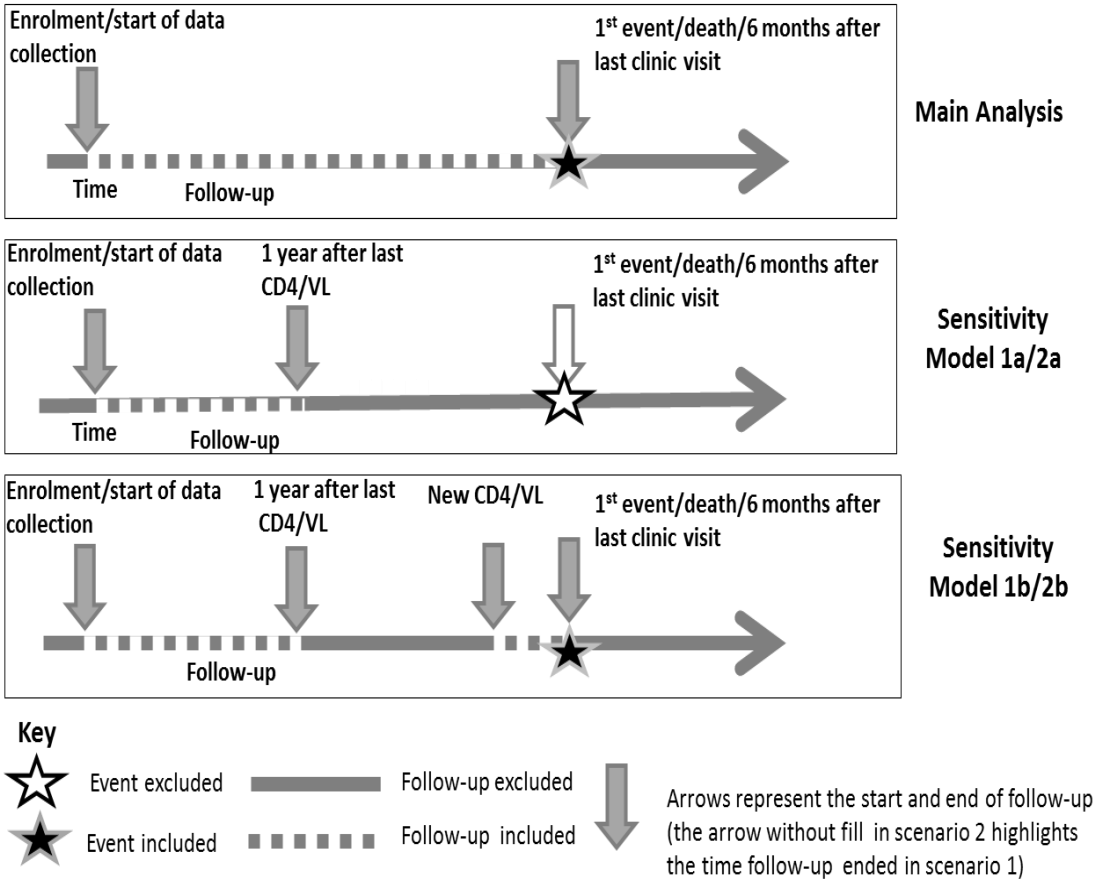


Figure 4.1: An illustration of how the main multivariable analysis (model J) and the sensitivity analyses (Models 1a-2b) differed with respect to follow-up time and event inclusion (using data from a hypothetical individual)

4.5.3 Sensitivity model 3: restricting the analysis to ≥01/01/2005

After the beginning of 2005 the association between HIV, ART and SNAEs started to be recognised in the literature and so recording of SNAEs may have improved. The final model from the main analysis (Model A) was refitted to a restricted dataset where earliest permitted entry to follow-up was 01/01/2005. Follow-up accrued and events occurring before 2005 were excluded.

4.5.4 Sensitivity model 4: changing the first possible date of follow-up entry to the day after the first recorded event for that cohort

Due to some uncertainty as to when systematic data capture began for each cohort, models B-J from the main analysis were refitted to an amended dataset, in which the start of follow-up for each individual began the day after the first event of that type was recorded by their cohort or at their enrolment, whichever happened later.

4.5.5 Sensitivity models 5 and 6: using multivariable fractional polynomials (mfp) to examine non-linear association [280]

Many associations between exposures and outcomes are non-linear, but regression modelling of continuous variables assumes that they are (or log-linear in the case of Poisson) [281]. To address this issue of non-linearity researchers often apply a limited number of established transformations (for example log transformations are commonly applied to HIV viral load) or use categorical variables [282]. The advantage of fractional polynomials is that they are very flexible and modelling with them elucidates the optimal transformation for a given variable by exploring many possible transformations [281]. I wanted to explore the validity of assuming linearity, with respect to the associations between continuous covariates and my outcomes (or which transformation would be optimal) and so I utilised this method.

In sensitivity model 5 variables retained in the final model (model J) of the main analysis were refitted choosing the best mfp model. The aim of this sensitivity analysis was to explore whether any transformations improved the fit of the final model from the main analysis.

In sensitivity model 6, mfp for all continuous variables was combined with backwards elimination for all variables ($p<0.1$, except for duration of HIV infection where p was set to 1). All available covariates were initially included unless they were collinear/strongly associated with each other (see Section 4.4.2.2.ii for details of how variables were selected if collinearity/strong association between variables was present).

The STATA mfp command, which works through a number of cycles, was used to fit both models. In brief, STATA cycles through all continuous covariates (from most to least statistically significant based on their linear effect) to determine the best-fitting fractional polynomial (FP) transformation for each (this is done in a 3-step approach outlined below). Further iterations assess improved selection of covariates and alternative transformations until convergence is achieved [283].

The three steps taken for each covariate in turn are as follows:

Step one: The test for Inclusion

Thirty-six separate fractional polynomial transformations (known as FP2s) of the selected variable were undertaken. These transformations are described algebraically by $\beta_1 X^{p1}+\beta_2 X^{p2}$ where p^1 and p^2 take the values shown in Table 4.5.

Table 4.5: Values p^1 and p^2 can take in mfp analyses

Value	Transformation
-2	$1/X^2$
-1	$1/X$
-0.5	$1/\sqrt{X}$
0	Log transformation
0.5	$\sqrt{X}$
1	X
2	X^2
3	X^3

When $p^1=1$ and $p^2=2$ this is equivalent to a quadratic regression and when $p^1=p^2$ this is called a repeated power model and takes the form $\beta_1 X^p+\beta_2 X^p \ln X$ [284].

For Sensitivity Model 5 the best fitting FP2 was retained for the next step. For Sensitivity Model 6 the best fitting FP2 was compared to the multivariable model without the variable and if the variable was not statistically significant ($p<0.1$ was used in my analyses) then the variable was dropped, otherwise the best fitting FP2 was retained for the next step.

Step two: The test for non-linearity

A model containing the best FP2 transformation was compared to a model where the variable was included in its linear form. If the transformation provided no statistically significant improvement in the model ($p<0.05$) then the linear form of the variable was kept, otherwise the transformed variable was retained for the next step.

Step three: The test for simplicity

The best fitting FP2 transformation of the variable was compared to the best fitting of eight non-fractional (FP1) transformations. FP1 transformations took the form X^p where p took each of the eight values shown in Table 4.4. The best fitting transformation (FP2 or FP1) was retained.

4.5.6 Sensitivity model 7: Restricting the analysis to cohorts with well validated MI cases

For MIs, an additional sensitivity analysis (model 7) was undertaken, restricted to those cohorts with well-validated cases which used standardised case definitions and used end-point review (FHDH, ICONA, and the Austrian HIV Cohort Study). A lack of robust case definitions for analyses involving MIs may introduce significant bias in the presence of a high proportion of false positives and negatives [285]. The final model from the main analysis (Model J) was refitted to the cohorts with well-validated cases.

5 Chapter 5: Patient Characteristics

5.1.Introduction

In this chapter I provide information on the characteristics of the sub-cohort of CASCADE participants who contributed data to my SNAE analyses. This aided my understanding of my study population and enabled me to assess the effect of possible sources of bias.

I examine (both overall and by cohort) numbers included (Section 5.2) and characteristics of these individuals (Section 5.3 I then assess the likely representativeness of my sample by comparing it to national data on the HIV-positive population in the UK and France (Section 5.4). This is followed by a chapter summary (Section 5.5).

5.2.Inclusion criteria

All patients who contributed any follow-up time to any of the events under investigation (fractures, MI or DM) were included. Some patient characteristics reported in this chapter are dependent on follow-up time, and left and right censoring dates varied by event type (see Chapter 4: Methods: Section 4.3). Follow-up for the purposes of this chapter started at the earliest date a patient entered follow-up and finished at the last date they left follow-up across all event types. Individuals excluded from contributing to any of the analyses and reasons for their exclusion are shown in Figure 5.1.

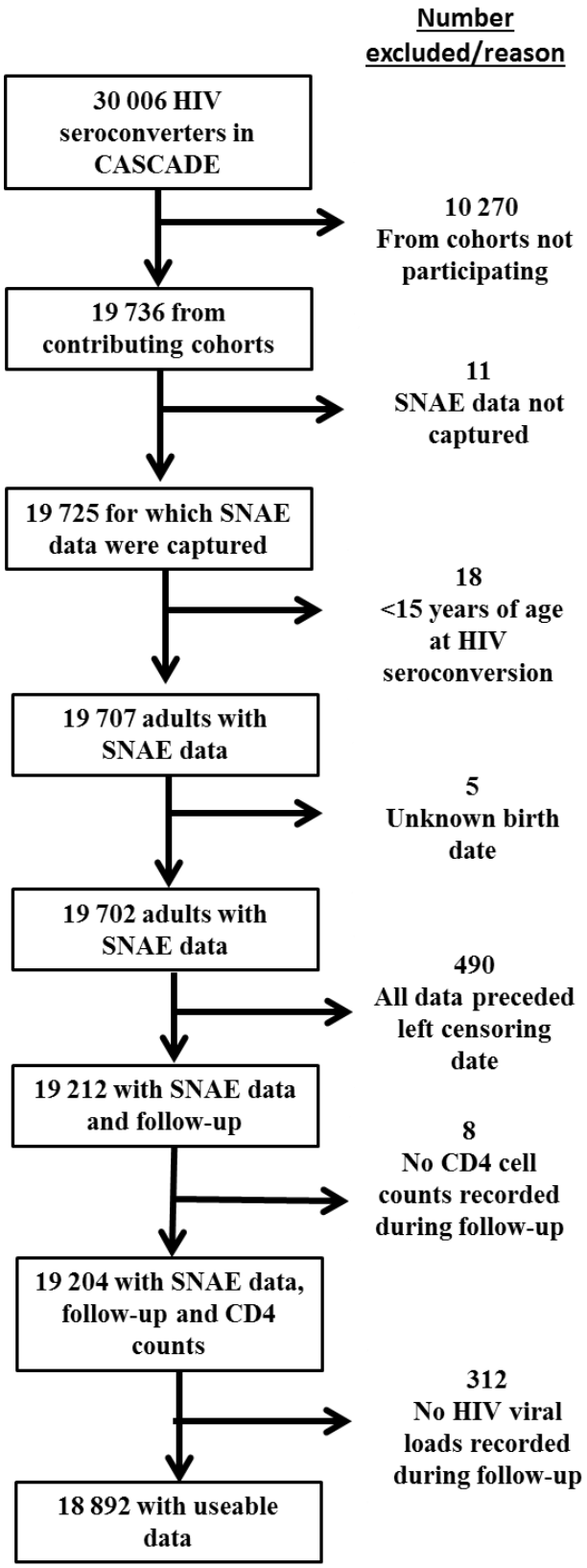


Figure 5.1: Flow chart showing the number of individuals included across all SNAE analyses and numbers and reasons for exclusion

5.3.Characteristics of my study sample

Table 5.1 summarises key characteristics of individuals included in my study. The median age at seroconversion was 32 years with males being slightly older than females. Data on ethnicity were missing for 60% (11 414/18 892) of individuals. This included all individuals from the largest contributing cohort (FHDH) where collection of these data is prohibited by national law. Therefore I used region of origin as a proxy for ethnicity where ethnicity was missing (see Chapter 4: Methods: Section 4.4.1). Transmission in men was predominantly through MSM and in females was largely through heterosexual sex. The FHDH contributed the largest proportion of individuals, 56.4% (10 651/18 892) of the total.

The validity of assumptions made about ethnicity using region of origin where ethnicity data were missing (see Chapter 4: Methods: Section 4.4 for these assumptions) was then assessed (Table 5.2). Agreement between ethnicity and assumed ethnicity (based on region of origin) was explored for 26.1% (4 938/18 892) of individuals for whom information on both variables was available.

For the 301 participants with black ethnicity, 74.8% (225/301) were categorised correctly as such based on the region of origin (Table 5.2). Similarly, 97.0% (4 338/4 472) of those with white ethnicity, and 80.6% (133/165) of those with other ethnicities were correctly categorised as such based on their region of origin. Cohen's Kappa coefficient for agreement between ethnicity and assumed ethnicity (based on region of origin) was 0.74 suggesting "good" agreement [286].

Table 5.1: Summary characteristics of individuals enrolled in CASCADE contributing data to any SNAE analysis (stratified by sex)

	Male (n=15 159)	Female (n=3 733)	Total (N=18 892)
	n (% of total excluding missing) or median [IQR]		
Seroconversion (age in years)	32.2 [26.8-39.1]	29.0 [24.1-35.7]	31.6 [26.2-38.6]
Ethnicity			
Black	244 (3.8)	152 (14.9)	396 (5.3)
White	6 026(93.4)	848 (83.0)	6 874 (91.9)
Other	186 (2.8)	22 (2.1)	208 (2.8)
Missing	8 703	2 711	11 414
Assumed ethnicity‡			
Black	881 (5.9)	991 (26.9)	1 872 (10.1)
White	13 780(92.3)	2 635(71.6)	16 415 (88.2)
Other	264 (1.8)	54 (1.5)	318 (1.7)
Missing	234	53	287
Mode of infection			
MSM	11 672(79.4)	2 (0.1)	11 674 (64.0)
PWID	699 (4.8)	408 (11.5)	1 107 (6.1)
MSW	2 027(13.8)	3046(86.1)	5 073 (27.8)
Other	293 (2.0)	83 (2.4)	376 (2.1)
Missing	468	194	662
Cohort			
FHDH (France)	8 036(53.0)	2 615(70.1)	10 651 (56.4)
AHIVCOS(Austria)	317 (2.1)	82 (2.2)	399 (2.1)
ICONA (Italy)	1 463(9.7)	398 (10.7)	1861 (9.9)
NOR (Norway)	433 (2.9)	69 (1.9)	502 (2.7)
PRIMO (France)	1 142(7.5)	200 (5.4)	1 342 (7.1)
SEROCO (France)	368 (2.4)	105 (2.8)	473 (2.5)
UKR (UK)	2 540(16.8)	195 (5.2)	2 735 (14.5)
CoRIS (Spain)	386 (2.6)	30 (0.8)	416 (2.2)
AMACS (Greece)	290 (1.9)	12 (0.3)	302 (1.6)
SAL (Canada)	184 (1.2)	27 (0.7)	211 (1.1)

MSM-men who have sex with men MSW-Men who have sex with women PWID-injecting drug use ‡ Based on region for those with missing ethnicity see Table 5.2

Table 5.2: Testing the validity of assumptions made about ethnicity based on region of origin*

Ethnicity	Region of Origin and assumed ethnicity			
	African/Caribbean ("Black")	Europe/North America/Australasia ("White")	Other ("Other")	Total
Black	225	65	11	301
White	53	4338	81	4472
Other	17	15	133	165
Total	295	4418	225	4938

*Only those individuals for whom both ethnicity and region of origin were known were included in this table

Figure 5.2 shows the age distribution by sex at the start of follow-up for the 18 892 participants contributing data to any SNAE analysis. There were 42.7% (8 073/18 892) persons aged 24-33 years and 31.4% (5 941/18 892) aged 34-43 years, with these two age groups making up 74.2% (14 014/18 892) of all those included in the analyses. Younger age groups tended to include a greater proportion of females; 32.7% (601/1 840) of individuals who started follow-up at <23 years of age were female, whilst just 13.1% (24/183) of those entering at ≥64 years were.

Table 5.3 summarises the characteristics of people included in the SNAE analyses stratified by cohort. The FHDH dominated the dataset contributing more than half of individuals and follow-up time (56% of total). Median follow-up time per patient (data not shown) was 5.6 [IQR, 2.4-10.8] years overall, but it varied by cohort. CoRIS patients contributed the shortest duration with a median of 1.9 [1.0-3.0] years and individuals from SEROCO contributed the most per person with a median of 10.3 [4.6-18.8] years. Overall 19.8% (3 733/18 892) of individuals were female but the percentage varied from 4.0% (12/302) for AMACS to 24.6% (2 615/10 651) for the FHDH. There was less variability in age at HIV seroconversion, with median values for most cohorts close to the overall median of 31.6 years. AMACS and SEROCO patients were the youngest at seroconversion, with medians of <30 years and PRIMO patients were the oldest with a median of 35.1 years. The majority of patients in all cohorts were infected through MSM, but the proportion infected through injecting drug use varied more widely from 0.9% (2/211) in the Southern Alberta clinic (the only non-European cohort) to 21.7% (392/1 861) for ICONA.

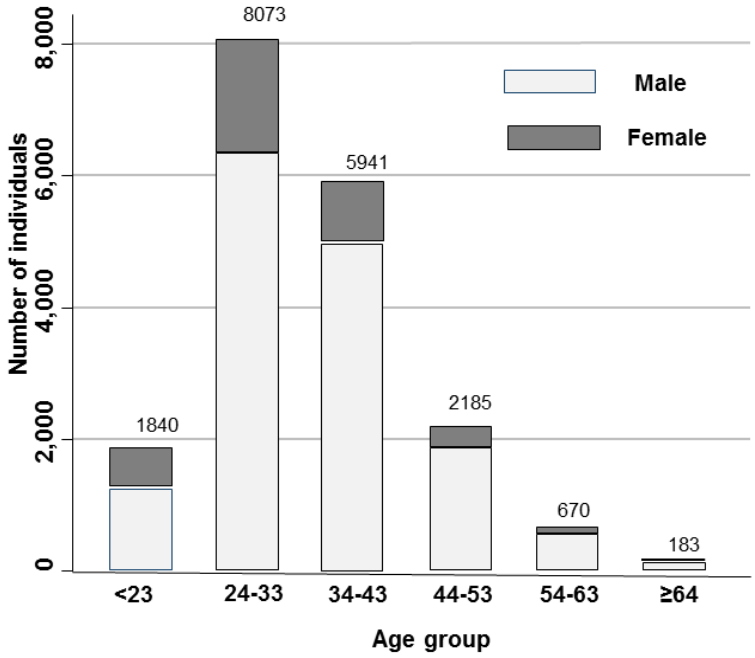


Figure 5.2: Age and sex distribution at start of follow-up for individuals contributing data to any SNAE analysis (N=18 892)

Approximately 10% of individuals enrolled in the UKR and CoRIS were heterosexually infected in contrast to >30% of those in the Austrian HIV Cohort and FHDH (Table 5.3). Mode of HIV infection was missing for 3.5% (662/18 892) of individuals. Overall, 86.9% were of white ethnicity/assumed ethnicity, but this varied from 99.0% (294/302) for AMACS to 78.6% (326/416) for CoRIS. Data were missing on both ethnicity and region in 1.6% (287/18 892).

Table 5.3: Summary characteristics of participants contributing follow-up to any SNAE analyses stratified by cohort

Cohort (N)	PYFU[¥]	Female	Seroconversion age (years)	Mode of Infection								Assumed ethnicity[#]						
				MSM		PWID		Heterosexual		Other		Missing	Black		White		Other		Missing
	n (%) or median [IQR]																	
FHDH (10651)	74 129	2 615 (24.6)	31.5 [26.1-38.5]	6044	(59.4)	458	(4.5)	3514	(34.5)	164	(1.6)	471	1383	(13.0)	9 175	(86.1)	93	(0.9)	0
AHIVCOS (399)	3 094	82 (20.6)	33.2 [26.5-40.6]	205	(51.9)	62	(15.7)	125	(31.7)	3	(0.8)	4	13	(3.3)	380	(95.5)	5	(1.3)	1
ICONA (1 861)	10 009	398 (21.4)	31.1 [26.3-37.6]	876	(48.4)	392	(21.7)	512	(28.3)	31	(1,7)	50	41	(2.2)	1 739	(93.4)	81	(4.4)	0
NOR (502)	4 596	69 (13.8)	33.1 [27.8-41.8]	329	(66.6)	42	(8.5)	120	(24.3)	3	(0.6)	8	23	(4.6)	457	(91.2)	21	(4.2)	1
PRIMO (1 342)	7 433	200 (14.9)	35.1 [29.0-43.0]	935	(73.5)	3	(0.2)	323	(25.4)	11	(0.9)	70	159	(11.9)	1 168	(87.2)	12	(0.9)	3
SEROCO (473)	5 309	105 (22.2)	28.4 [24.4-34.4]	303	(66.2)	14	(3.1)	118	(25.8)	23	(5.0)	15	11	(2.3)	457	(96.8)	4	(0.9)	1
UKR (2 735)	22 746	195 (7.1)	31.0 [25.9-37.4]	2325	(85.7)	107	(3.9)	245	(9.0)	37	(1.4)	21	149	(6.1)	2 229	(90.5)	84	(3.4)	273
CoRIS (416)	923	30 (7.2)	32.2 [26.7-37.8]	349	(84.1)	18	(4.3)	45	(10.8)	3	(0.7)	1	85	(20.5)	326	(78.6)	4	(1.0)	1
AMACS (302)	2 717	12 (4.0)	29.4 [25.2-29.4]	176	(62.6)	9	(3.2)	34	(12.1)	62	(22.1)	21	3	(1.0)	294	(99.0)	0	(0.0)	5
SAL (211)	1 118	27 (12.8)	31.0 [25.3-39.6]	132	(62.6)	2	(0.9)	37	(17.5)	39	(18.5)	1	5	(2.4)	190	(90.0)	14	(6.6)	2
Total (18892)	132 074	3 733 (19.8)	31.6 [26.2-38.6]	11674	(62.8)	1107	(5.9)	5073	(26.9)	376	(2.0)	662	1872	(15.7)	16 415	(86.9)	318	(1.7)	287

[#] Region was used as a proxy for ethnicity where ethnicity was missing.[¥] Person-years of follow-up

The last-patient last-visit dates by cohort for those included in any SNAE analysis are shown in Table 5.4. Last-patient last-visit dates were defined as the last recorded date of a CD4 cell count, HIV viral load or ART treatment change.

Table 5.4: Cohort last-patient last-visit dates for individuals included in any SNAE analysis

Cohort	Last patient last visit date
FHDH	11/09/2012
AHIVCOS	26/03/2013
ICONA	30/01/2013
NOR	12/02/2013
PRIMO	27/02/2013
SEROCO	9/12/2009
UKR	14/01/2013
CoRIS	28/10/2011
AMACS	14/02/2013
SAL	11/02/2013

The last-patient last-visit date was 9/12/2009 for SEROCO and 28/10/2011 for CoRIS, otherwise dates were within around 6 months of each other, ranging from 11/09/2012 to 26/03/2013.

Figure 5.3 shows follow-up status on the day of the last-patient last-visit date for each cohort. Individuals were classified as being in one of three groups, active follow-up, lost to follow-up or deceased. To be assigned to the active follow-up group an individual must have had a CD4 cell count, HIV viral load, ART treatment change or known “alive” status recorded in the year prior to their cohort’s last-patient last-visit date and no record of death before that date. Those considered lost to follow-up were individuals not meeting these criteria, but with no record of death.

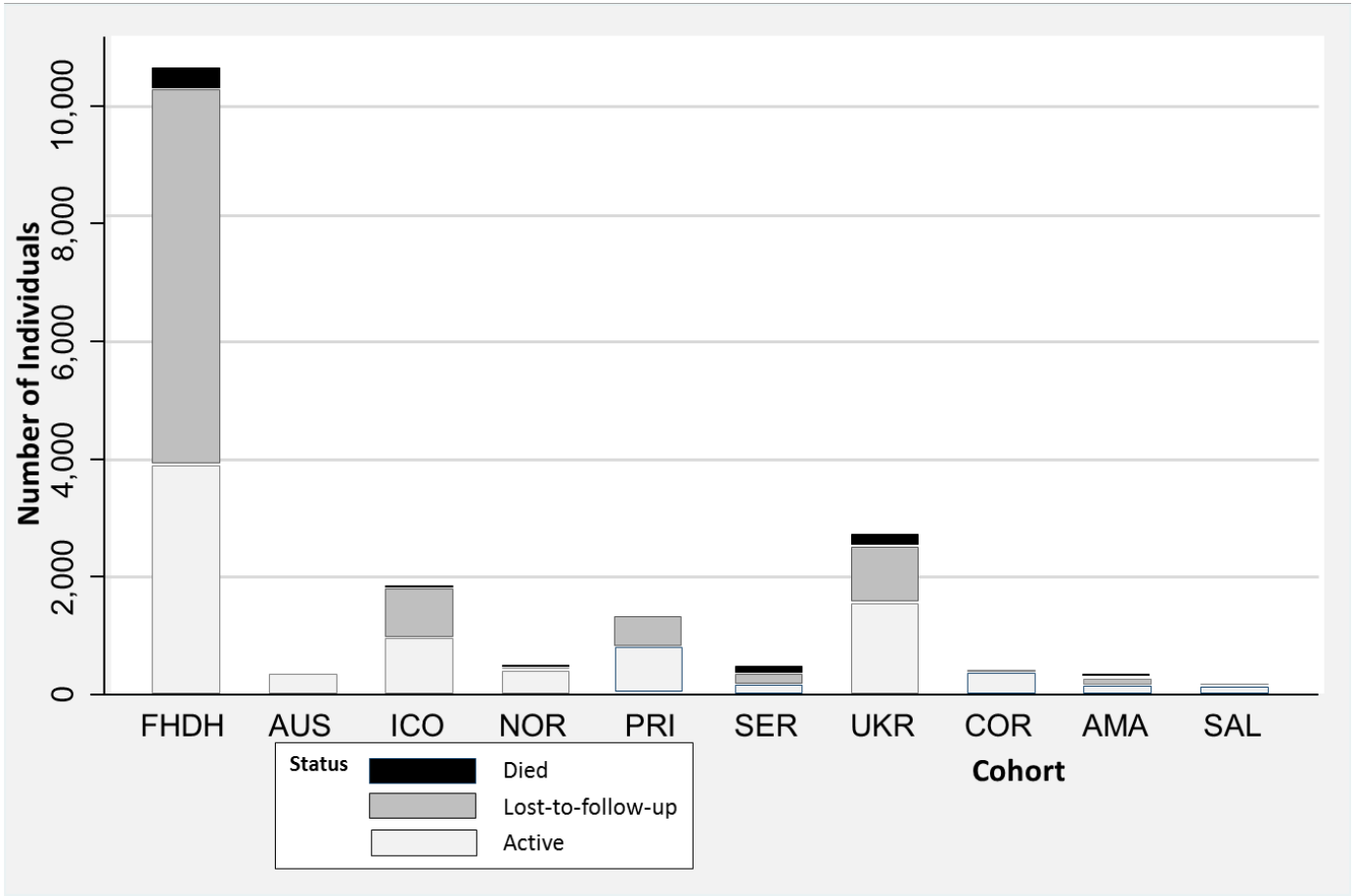
In total 46.3% (8 752/18 892) of individuals were classified as still in active follow-up, 48.8% (9 210/18 892) were classified as lost to follow-up and 4.9% (930/18 892) were known to have died. Follow-up status was highly variable between cohorts. It is notable that 60.0% (6 386/10 651) of those enrolled in FHDH and 35.5% (970/2 735) of those in the UK Register (970/2 735) were classified as lost to follow-up in contrast to 6.5% (26/399) from the Austrian HIV cohort study and 8.4% (42/502) from the Oslo and Ulleval hospital cohort. The proportion of deaths also varied markedly by cohort. SEROCO recorded 27.9% (132/473) of individuals as having died during follow-up, with AMACS recording 11.9%

(36/302) and the Oslo and Ulleval hospital cohort recording 11.2% (56/502). This contrasted with FHDH, ICONA, PRIMO and CoRIS who all had <4% of individuals recorded as dying during follow-up.

Table 5.5 highlights characteristics of individuals contributing data to any SNAE analyses at different time points during follow-up: follow-up start (split by prior ART exposure), ART start (for those who started ART during follow-up on any regimen including sub-optimal ones) and at the end of follow-up. The majority of individuals started follow-up ART naïve (82%) and of these 76% started ART during follow-up.

There were some differences in the characteristics of those ART-naïve at follow-up start compared to those who were ART-experienced. ART-experienced individuals were more likely to start follow-up: in earlier calendar years, at an older age, with a prior AIDS diagnosis, HCV seropositive, with longer duration of HIV infection, with a higher probability of having a CD4 cell count and viral load measurement recorded and with a higher recent CD4 cell count.

From ART start to the end of follow-up the percentage of individuals on: NNRTI based regimens decreased, PI regimens remained similar, integrase inhibitors increased and fusion inhibitors increased. By the end of follow-up individuals had spent an average of nearly 3 years on ART.



"Active" follow-up was defined as having had a CD4 cell count, HIV viral load, ART treatment change or known alive status recorded in the year prior to the cohort's last-patient last-visit date in individuals who were not recorded as having died.

Figure 5.3: Status of all individuals included in any SNAE analysis at cohort last-patient-last-visit date (2013 dataset)

Table 5.5: Characteristics of HIV-positive individuals who contributed follow-up to any SNAE analysis at key time points

	Start of follow-up		ART start*	End of follow-up
	ART naïve (n=15 512)	Prior ART (n=3 380)	(n=11 722)	(n=18 892)
	n (%) or median [IQR] or IRR (95%CI)			
Calendar time (mm/yy)	08/04 [06/98-09/08]	01/00 [06/98-08/06]	04/06 [11/99-10/09]	01/12 [05/09-09/12]
Current age (years)	33.0 [27.6-39.9]	34.9 [29.8-41.7]	35.5 [29.9-42.6]	41.1 [34.1-48.1]
Years since HIV seroconversion	0.6 [0.2-1.3]	1.5 [0.29-5.1]	2.3 [1.1-4.8]	7.0 [3.4-12.6]
No of CD4 cell counts per PYFU	-	-	2.9 (2.9-2.9)	3.0 (3.0-3.0)
Individuals with CD4 value at time point (n)#	11 339 (73.1)	2 987 (88.4)	11500 (98.1)	18 892 (100.0)
CD4 cell count at time point (cells/µL)	497 [354-674]	364 [243-540]	330 [237-453]	558 [405-741]
Prior AIDS	169 (1.1)	279 (8.3)	591 (5.0)	1 898 (10.1)
No of viral load measurements per PYFU¥	-	-	2.5 (2.5-2.5)	2.7 (2.7-2.7)
Individuals with HIV viral load value (n)#	7 456 (48.1)	1 862 (55.1)	10 051 (85.7)	18 892 (100.0)
HIV viral load at time point (log_{10} copies/ml)	4.4 [3.7-4.9]	4.1[2.4-5.0]	4.7 [4.0-5.2]	1.4 [1.3-3.5]
No. with HCV test recorded during FU	-	-	3 420 (29.2)	7995 (42.3)
No of HCV tests per 100 PYFU	-	-	24.4 (23.7-25.0)	19.1 (18.9-19.4)
Number of HCV sero-positive patients~	832 (5.4)	232 (6.9)	901 (7.7)	2 004 (10.6)
ART status at time point				
Naive	15 512 (100.0)	-	-	3 790 (20.1)
Off ART	-	176 (5.2)	-	745 (4.0)
Sub-optimal ART	-	1 225 (36.2)	3 674 (31.3)	2 552 (13.5)
cART≠	-	1 979 (58.6)	8 048 (68.7)	11 805 (62.5)
Current regimen includes[1]:	-			
NNRTI	-	600 (3.2)	3 749 (33.0)	5 676 (30.0)
PI	-	1 390 (41.1)	3 848 (32.8)	6 031 (31.9)
Fusion Inhibitor/Integrase Inhibitor	-	112 (3.5)	353 (3.0)	1 517 (8.0)

ART-antiretroviral therapy NNRTI-Non-nucleotide reverse transcriptase inhibitor PI-protease inhibitor ≠cART-at least 3 drugs of 2 or more classes #Current CD4/VL were carried forward from one measurement to the next. ~Individuals classified as HCV sero-positive are those who ever had a prior recorded positive test (see section 4.4.1 for more details). *Excludes those who had started ART prior to follow-up start and also excludes those who did not start ART during follow-up, but includes sub-optimal regimens PYFU-Person-years of follow-up [1]Does not add up to 100% as individuals are usually on a number of ART drugs simultaneously.

Table 5.6 summarises characteristics at the end of follow-up for PLHIV contributing to any SNAE analysis, stratified by cohort. Median age at follow-up end varied by cohort; those in CoRIS were the youngest (34 years) and those in the Oslo and Ulleval hospital cohort (NOR) the oldest (45 years). Those enrolled in CoRIS had been exposed to HIV for the shortest (median) time by the end of follow-up (3 years). By contrast SEROCO patients had had a median of 11 years since HIV seroconversion and those in the UK Register 10 years.

Overall, by the end of follow-up recorded rates of CD4 measurement were 3.0 per year, but varied by cohort being higher in AHIVCOS and lower in SEROCCO, AMACS and FHDH.

Table 5.7 shows ART status at follow-up end stratified by cohort. ART status varied markedly by cohort, with 66% of CoRIS patients being ART naïve at follow-up end compared to 31% in ICONA and just 12% for the Austrian HIV cohort.

Sub-optimal ART regimens were reported for around 20% of SEROCO and ICONA patients compared to ≤3% in AHIVCOS, NOR and SAL. AHIVCOS and NOR also had the highest percentages of patients on cART at the end of follow-up (82% and 77% respectively). In those ever starting ART, median time spent on ART by the end of follow-up varied from around one year in CoRIS to over 8 years in SEROCO.

Table 5.6: Summary characteristics of HIV-positive individuals contributing follow-up to any SNAE analysis stratified by cohort

	FHDH (10 651)	AHIVCOS (399)	ICONA (1 861)	NOR (502)	PRIMO (1 342)	SEROCO (473)	UKR (2 735)	CoRIS (416)	AMACS (302)	SAL (211)
	n (%) or median [IQR] or rate (95%CI)									
Age (years)	40.9 [34.1-47.8]	42.0 [34.0-50.2]	40.3 [33.8-46.5]	45.1 [37.8-52.0]	41.1 [34.3-49.0]	42.5 [34.1-48.9]	43.0 [36.1-49.5]	34.3 [29.9-40.6]	39.7 [33-48.2]	38.6 [31-48.1]
Years since seroconversion	7.1 [3.7-12.5]	6.7 [3.5-12.0]	6.1 [3.0-11.9]	8.8 [4.3-14.7]	4.5 [2.3-8.1]	11.0 [5.6-19.5]	9.6 [4.6-16.2]	2.7 [1.7-4.0]	7.2 [3.1-14.2]	5.4 [2.4-9.2]
CD4 counts/PYFU¥	2.9 (2.9-2.9)	3.6 (3.5-3.6)	3.1 (3.1-3.2)	3.1 (3.0-3.1)	3.2 (3.2-3.3)	2.4 (2.4-2.5)	3.2 (3.2-3.3)	3.2 (3.1-3.4)	2.6 (2.6-2.7)	3.0 (2.9-3.1)
CD4# (cells/µL)	552 [403-735]	645 [479-803]	550 [407-743]	543 [390-727]	638 [481-811]	416 [192-604]	562 [404-740]	541 [415-717]	582 [366-791]	462 [323-641]
Prior AIDS	1 061 (10.0)	35 (8.8)	111 (6.0)	60 (12.0)	34 (2.5)	144 (30.4)	386 (14.1)	13 (3.1)	26 (8.6)	28 (13.3)
Viral loads/PYFU¥	2.5 (2.5-2.6)	3.5 (3.5-3.6)	2.9 (2.9-3.0)	2.7 (2.7-2.8)	3.4 (3.3-3.4)	1.7 (1.6-1.7)	3.3 (3.2-3.3)	3.1 (3.0-3.3)	2.0 (1.9-2.0)	3.1 (3.0-3.3)
VL#(log_{10}cps/ml)	1.4 [1.4-3.3]	1.3 [1.0-1.6]	2.3 [1.4-4.0]	1.0 [1.0-2.8]	1.3 [1.0-2.3]	3.0 [1.3-4.5]	1.4 [1.3-3.3]	3.9 [1.6-4.6]	1.4 [1.4-3.8]	1.3 [1.3-3.1]
Ever HCV test	U	389 (97.5)	1617 (86.9)	494 (98.4)	1252 (93.3)	220 (46.5)	2 507 (91.7)	397 (95.4)	296 (98.0)	210 (99.5)
No. HCV sero-positive~	612 (5.7)	60 (15.0)	521 (28.0)	59 (11.8)	43 (3.2)	67 (14.2)	152 (5.6)	28 (6.7)	71 (23.5)	35 (16.6)

#Current CD4/VL were carried forward from one measurement to the next. ~Individuals classified as HCV sero-positive were those who ever had a prior recorded positive test (see section 4.4.1 for more details). U-unknown, because FHDH only sent information on positive test results & not tests that had been found to be negative ® PYFU-Person-years of follow-up

Table 5.7: ART status of HIV-positive individuals contributing follow-up to any SNAE analysis at follow-up end stratified by cohort

	FHDH (10 651)	AHIVCOS (399)	ICONA (1 861)	NOR (502)	PRIMO (1 342)	SEROCO (473)	UKR (2 735)	CoRIS (416)	AMACS (302)	SAL (211)
	n (%) or median [IQR]									
ART status										
Naive	1 899 (17.8)	46 (11.5)	571 (30.7)	70 (13.9)	195 (14.5)	107 (22.6)	526 (19.2)	273 (65.6)	69 (22.9)	34 (16.1)
Off ART	155 (1.5)	16 (4.0)	114 [6.1)	32 (6.4)	101 (7.5)	57 (12.1)	207 (7.6)	4 (1.0)	15 (5.0)	44 (20.9)
ART	1 784 (16.8)	11 (2.8)	365 (19.3)	15 (3.0)	67 (5.0)	99 (20.9)	170 (6.2)	13 (3.1)	22 (7.3)	6 (2.8)
cART≠	6 813 (64.0)	326 (81.7)	811 (43.6)	385 (76.7)	979 (73.0)	210 (44.4)	1 832 (67.0)	126 (30.3)	196 (64.9)	127 (60.2)
Years on ART* (if ever started)	4.28 [1.67-9.52]	3.60 [1.75-7.39]	2.82 [0.98-6.81]	4.06 [1.88-8.81]	2.74 [1.29-5.40]	8.11 [2.9-12.44]	5.70 [2.53-10.81]	1.25 [0.47-2.31]	5.21 [1.69-10.79]	2.7 [0.9-5.12]
ART class										
NNRTI-based	3 097 (29.1)	147 (36.8)	425 (22.8)	163 (32.5)	392 (29.2)	98 (20.7)	1 132 (41.4)	65 (15.6)	94 (31.1)	63 (29.9)
PI-based	3644 (34.2)	140 (35.1)	450 (24.2)	200 (39.8)	492 (36.7)	123(26.0)	811 (29.7)	39 (9.4)	96 (31.8)	36 (17.1)
FI/II-based~	930 (8.7)	45 (11.3)	96 (5.2)	29 (5.8)	168 (12.5)	35 (7.4)	173 (6.3)	14 (3.4)	10 (3.3)	17 (8.1)

* This includes any time on ART before follow-up started ~Fusion or Integrase Inhibitor ≠cART-at least 3 drugs of 2 or more classes

Of the 11 114 individuals excluded from all SNAE analyses, 844 were from cohorts which provided data to CASCADE on SNAEs. All cohorts had excluded individuals, but 63.7% (538/844) of all exclusions were from the UK Register. When the characteristics of these 844 excluded individuals were compared to the 18 892 included individuals, the most striking difference between the two groups was the median date of HIV seroconversion, which was markedly earlier in those excluded (1990 vs. 2002). Early HIV seroconverters were more likely to have had no CD4 cell counts or HIV viral load measurements recorded during follow-up and/or their follow-up was more likely to have ended before systematic data collection of SNAEs commenced, hence their exclusion from the analyses.

To allow a more meaningful comparison of those included and excluded (to reduce the effects due to early calendar year), Table 5.8 restricts the comparison to those who seroconverted ≥01/01/1996. Using this criterion, there were 262 individuals excluded (of the 844 excluded in total) and 14 242 individuals included (of 18 892 included in total) who seroconverted ≥01/01/1996.

The proportion of individuals excluded varied by cohort. FHDH patients were less likely to be in the excluded group and individuals from the UK Register were more likely to be so. This is due to differences in left censoring dates between cohorts. FHDH patients were included from cohort inception, whilst UK Register patients were included from the data after the first SNAE was recorded for the cohort, due to uncertainty when systematic data collection began (16/06/1998). Those excluded were also more likely to be of black or "other" ethnicity and MSM.

Table 5.8: Comparison of baseline characteristics of participants who seroconverted ≥01/01/1996 and were included and excluded from any SNAE analysis

N=14 504	Included (n=14 242)		Excluded (n=262)		p
Seroconversion (age in years)	32.8 [27.2-39.8]		32.1 [26.9 39.8]		0.8*
Ethnicity					
Black	345	(6.2)	15	(9.6)	0.02
White	5 017	(90.5)	132	(84.1)	
Other	180	(3.3)	10	(6.4)	
Missing	8 700		105		
Assumed Ethnicity					
Black	1 665	(11.8)	34	(13.3)	0.05
White	12 219	(86.3)	212	(82.8)	
Other	271	(1.9)	10	(3.9)	
Missing	87		6		
Mode of infection					
MSM	9 139	(66.8)	185	(75.8)	0.007
PWID	422	(3.1)	10	(4.1)	
Heterosexual	3 931	(28.7)	46	(18.9)	
Other	200	(1.5)	3	(1.2)	
Missing	550		18		
Cohort					
FHDH	8 143	(57.0)	73	(27.9)	<0.0001
AHIVCOS	356	(2.5)	9	(3.4)	
ICONA	1 360	(9.6)	15	(5.7)	
NOR	370	(2.6)	8	(3.1)	
PRIMO	1 342	(9.4)	49	(18.7)	
SEROCO	23	(0.2)	1	(0.4)	
UKR	1 822	(12.8)	78	(29.8)	
CoRIS	416	(2.9)	23	(8.8)	
AMACS	221	(1.6)	2	(0.8)	
SAL	189	(1.23)	4	(1.5)	

*Two-sample Mann-Whitney test which excluded five patients with a missing date of birth, all other p values use chi-squared MSM-men who have sex with men PWID-injecting drug use Clinic

5.4. Study population representativeness

I compared the characteristics of UK PLHIV included in my study to those reported by Public Health England (PHE) in its 2013 annual report.

The PHE data includes all HIV-positive adults treated in the UK. The UK Register, the only UK contributor to CASCADE, is a sero-incident cohort [287]. White/non-African born males who contracted HIV through the MSM route are over-represented in my sample compared to the PHE population (Table 5.9) [288]. This observed difference will be an underestimate (of up to 3%) as some individuals included in the PHE data will also be enrolled in the UK Register. Also of note, is that the PHE data only includes individuals who were alive and in treatment in the UK in 2013. In contrast, my UK Register data includes individuals in follow-up prior to 2013, some of whom may have died or moved abroad by 2013. Therefore the comparison is approximate. Characteristics of these two groups are compared in Table 5.9.

Table 5.9: A comparison of demographic characteristics of individuals enrolled in the UK Register (and included in my study) with all HIV-positive people treated in the UK in 2013

	Public Health England (2013 data)		UK Register (& included in my study)	
	N/%			
N	98 400*		2 735	
Male	65 600	67	2 540	93
Transmission route				
MSM	41 000	42	2 325	86
PWID	2 200	2	104	4
Heterosexual	53 000	54	245	9
Other	2 200	2	61	2
Born in Africa†	31 800	32	<233	<9

*This included an estimated 21 900 who were thought to be HIV-positive but undiagnosed †PHE data excludes PWID and MSM born in Africa, max. estimate for UK register as all non-white assumed to be born in Africa!

In the FHDH (which is known to be broadly representative of PLHIV in France) as a whole 67% were male and 34% MSM [289]. In contrast, 75% of FHDH patients included in my study (i.e. those with well-estimated seroconversion time) were male and 59% were MSM, suggesting preferential selection of MSM in my largest contributing cohort as well. The comparison, will again under-estimate the difference as those enrolled in FHDH and contributing to my study also contribute data to the FHDH as a whole.

5.5.Summary

Just over half of the entire CASCADE cohort was included in my analysis. The majority of individuals were excluded because they provided no data on outcomes (predominantly because it was not captured by their cohort). A small percentage (4%) were excluded because they had no new data after the left censoring date or because there were no CD4 or HIV viral load data recorded (3%) during follow-up.

Individuals from France dominated the dataset with (56%) of all those included from FHDH with additional French contributions from PRIMO and SEROCO. The UK was also a major contributor (15%). The population was largely white (88%), male (80%), started follow-up aged in their early 30s and contracted HIV through MSM (64%). White MSM were over-represented in my sample compared to the HIV-positive population of France and the UK.

At the start of follow-up the study population was largely ART-naïve and had recently HIV seroconverted. Most of these individuals started ART during follow-up. Median follow-up was a little over 5 years. The last-patient-visit date was in March 2013.

By follow-up end about 50% of subjects were classified as lost to follow-up, 5% had died, 11% were known to be HCV sero-positive and 10% had ever had an AIDS.

6 Chapter 6: Fracture Incidence

6.1 Introduction

Fracture rates are increased in those with HIV, but the reasons for this are unclear [119]. They may result from: some aspect of HIV infection affecting bone-turnover, ART exposure, HCV or other factors [188, 290]. This chapter examines the association between HIV-related variables and fracture incidence before and after adjustment for established fracture predictors (where available).

In Section 6.2 I show which individuals were included in my analysis and provide numbers and reasons for exclusion. I then go on to describe baseline characteristics (section 6.3) before presenting my unadjusted results (section 6.4). In section 6.5 I explore the effect of adjustment. Section 6.6 outlines the results of my sensitivity analyses and is followed by a summary of my findings (Section 6.7). I then go on to compare my study characteristics and findings with those of other studies (Section 6.8).

6.2 Inclusion

In total, 13 600 individuals with 300 first fractures from five cohorts were potentially eligible for inclusion in the analysis. Figure 6.1 indicates the reasons for exclusion and the number of individuals and events excluded. In total, 34 events (11%) and 357 (3%) individuals were excluded from the analysis (Figure 6.1). 13 243 people with a total follow-up of 89 470 person years (median 5.3 [2.5-10.1] years) were included; there were 266 first fractures, equivalent to an event rate per 1000 PYFU of 3.0 (95%CI, 2.6-3.4).

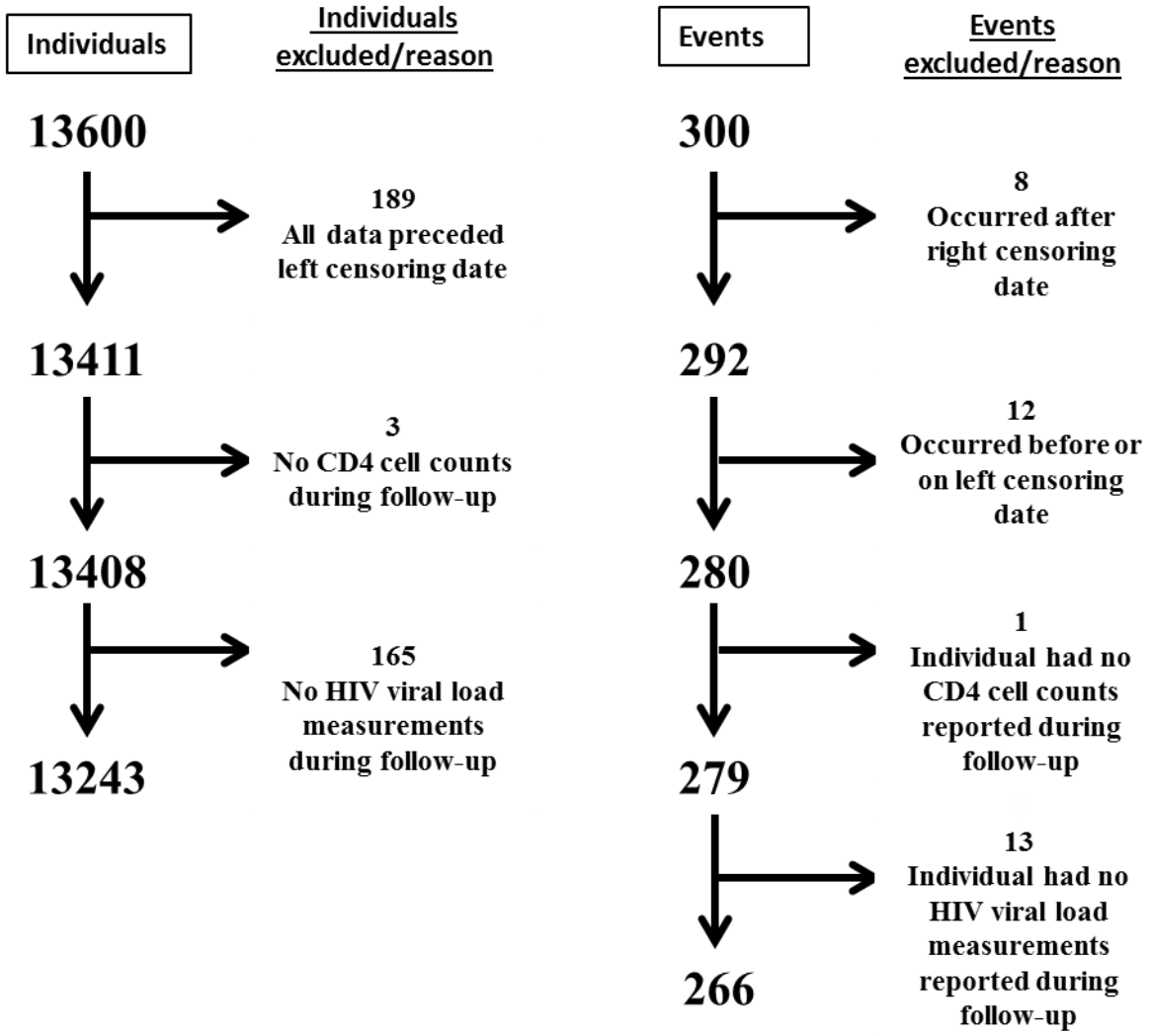


Figure 6.1: Numbers of HIV-positive individuals and events included in the fractures analysis with numbers and reasons for exclusion

6.3 Baseline characteristics

Table 6.1 shows baseline characteristics of all individuals included in my analysis of fractures. The median age at seroconversion was 31.8 years with males being slightly older than females. The FHDH contributed the majority of individuals (80.2%). Transmission in men was predominantly MSM and in females was largely through heterosexual sex. A greater percentage of females were of black ethnicity (28.3%) than males (5.8%).

Table 6.1: Baseline characteristics of individuals in the analysis examining the association between duration/markers of HIV infection and fracture incidence, stratified by sex

	Male (n=10 198)		Female (n=3 045)		Total (N=13 243)	
	Median [IQR] or n (%≠)					
Seroconversion age (years)	32.7 [27.1-39.7]		29.0 [24.0-36.2]		31.8 [26.3-39.0]	
Assumed ethnicity[2]						
Black	585	(5.8)	862	(28.3)	1 447	(11.0)
White	9 474	(93.5)	2 118	(70.3)	11 592	(88.2)
Other	76	(0.7)	31	(1.0)	107	(0.8)
Missing	63		34		97	
Mode						
MSM#/bisexual	7 759	(79.1)	1	(0)	7 760	(61.2)
PWID~	318	(3.2)	234	(8.2)	552	(4.4)
Heterosexual	1 596	(16.3)	2566	(89.7)	4162	(32.8)
Other	141	(1.4)	61	(2.1)	202	(1.6)
Missing	384		183		567	
Cohort						
FHDH	8 018	(78.6)	2 608	(85.7)	10 626	(80.2)
AHIVCOS[3]	314	(3.1)	81	(2.7)	395	(3.0)
NOR[4]	357	(3.5)	53	(1.7)	410	(3.1)
PRIMO	1 141	(11.2)	198	(6.5)	1339	(10.1)
SEROCO	368	(3.6)	105	(3.4)	473	(3.6)

[1]Missing ethnicity also included prohibited. [2]Region of origin was used to determine likely ethnicity when ethnicity was missing. [3]AHIVCOS- Austrian HIV cohort [4]NOR-Oslo and Ulleval hospital cohort #MSM, men who have sex with men ~PWID intravenous drug users ≠Percentage of non-missing total

6.4 Univariable analysis

6.4.1 Time-fixed variables

In the univariable analysis of time-fixed variables fracture rates increased with increasing age at seroconversion. Individuals aged ≥60 years at seroconversion experienced 3.5-fold higher rates than individuals aged <20 years (Table 6.2). Women had a lower risk of fractures than men, although the test for heterogeneity was of borderline significance (p=0.07). Those of white ethnicity had a higher risk of fractures than those in the black or "other" groups. PWID were more likely to experience fractures than MSM or heterosexuals. Fracture rates varied substantially by cohort. In FHDH, the largest cohort, the fracture rate was lowest (2.28 per 1000 PYFU (1.96-2.66)). Fracture rates in PRIMO were similar to those in FHDH. Rates were much higher in The Austrian HIV cohort (17.42 per 1000 PYFU (13.13-23.14)) and in the Oslo and Ulleval hospital cohort (7.16 per 1000 PYFU (3.58-14.32)).

Table 6.2: The association between time-fixed factors and fracture incidence in HIV-positive individuals: Rates and univariable analysis

Time-fixed factor	Event No.	PYFU*	Rate/1000 PYFU (95%CI)	IRR (95%CI)	p[1]
Seroconversion age (years)					
<20	5	3 110	1.61 (0.67-3.86)	1	0.03
20-<40	193	69 009	2.80 (2.43-3.22)	1.74 (0.72-4.20)	
40-<60	61	16 141	3.78 (2.94-4.86)	2.35 (0.94-5.85)	
≥60	7	1 212	5.78 (2.75-12.2)	3.59 (1.14-11.3)	
Sex					
Male	211	66 564	3.17 (2.77-3.63)	1	0.07
Female	55	22 906	2.40 (1.84-3.13)	0.76 (0.56-1.02)	
Ethnicity					
White	97	14 985	6.47 (5.31-7.90)	1	0.2
Black	1	1 047	0.95 (0.13-6.78)	0.15 (0.02-1.06)	
Other	0	150	-	-	
Missing	168	73 288		0.35 (0.28-0.45)	
Assumed Ethnicity[2]					
Black	10	8415	1.19 (0.64-2.21)	1	0.007
White	255	79961	3.19 (2.82-3.61)	2.68 (1.43-5.05)	
Other	1	602	1.66 (0.23-11.8)	1.40 (0.18-10.9)	
Missing	0	493	-	-	
Mode					
MSM[#]	141	49 142	2.87 (2.43-3.38)	1	0.005
PWID~	28	5 525	5.07 (3.50-7.34)	1.77 (1.18-2.65)	
Heterosexual	80	29 701	2.69 (2.16-3.35)	0.94 (0.71-1.24)	
Other	10	1 781	5.61 (3.02-10.4)	1.96 (1.03-3.71)	
Missing	7	3 322	2.11 (1.00-4.42)	0.73 (0.34-1.57)	
Mode (binary)					
Non-PWID	231	80 624	2.87 (2.52-3.26)	1	0.004
PWID~	28	5 525	5.07 (3.50-7.34)	1.77 (1.19-2.62)	
Missing	7	3 322	2.11 (1.00-4.42)	0.74 (0.35-1.56)	
Cohort					
FHDH	167	73130	2.28 (1.96-2.66)	1	<0.0001
AHIVCOS[3]	48	2753	17.4 (13.1-23.1)	7.63 (5.54-10.5)	
NOR[4]	8	1117	7.16 (3.58-14.3)	3.14 (1.54-6.38)	
PRIMO	21	7354	2.86 (1.86-4.37)	1.25 (0.79-1.97)	
SEROCO	22	5118	4.30 (2.83-6.53)	1.88 (1.21-2.94)	

[#]MSM-Men who have sex with men ~PWID-Intravenous drug users [1]Test for heterogeneity excluding missing [2] Region of origin was used as a guide for ethnicity when ethnicity was missing. [3]AHIVCOS- Austrian HIV cohort [4]NOR-Oslo and Ulleval hospital cohort *PYFU-person-years of follow-up

6.4.2 Time-updated variables

Fracture rates increased with increasing duration of HIV infection (Table 6.3); rates nearly doubled in those with >12 years exposure versus those with ≤3 years. Fracture rates increased with increasing current age; those with current age >60 years experienced rates nearly five times those ≤30 years. There was no evidence of a statistically significant association between time-updated CD4 cell count and fractures. Fracture rates increased with decreasing nadir CD4 cell count. Fracture rates were highest in those with HIV viral loads <1000 copies/mL, with little variability in rates by viral load category in individuals with viral load ≥1000 copies/mL. Individuals on any current ART experienced higher rates of fractures compared to those not on current ART. Those receiving concurrent PI and TDF experienced the highest rates. Fracture rates increased with increasing duration of immune-suppression at all levels. There was evidence that rates of fractures were elevated (nearly two fold) in individuals who had had a prior AIDS event (more than two fold). Those with HCV seropositivity experienced more fractures compared to those not known to have ever been exposed to HCV. Fractures rates increased in later calendar periods.

Table 6.3 (Page 1 of 2): The association between time-updated factors and fracture incidence in PLHIV: rates and univariable analysis

Time-updated factor	Events (N)	PYFU*	Rate/1000 PYFU(95%CI)	IRR (95%CI)	p[1]
Duration of HIV Infection					
≤3 years	49	23 327	2.10 (1.59-2.78)	1	0.003
>3-6	59	22 965	2.57 (1.99-3.32)	1.22 (0.84-1.79)	
>6-9	53	16 594	3.19 (2.44-4.18)	1.52 (1.03-2.24)	
>9-12	43	11 340	3.79 (2.81-5.11)	1.81 (1.20-2.72)	
>12	62	15 245	4.07 (3.17-5.21)	1.94 (1.33-2.82)	
Current age					
≤30 years	19	14 895	1.28 (0.81-2.00)	1	<0.0001
>30-40	96	35 808	2.68 (2.19-3.27)	2.10 (1.28-3.44)	
>40-50	85	25 603	3.32 (2.68-4.11)	2.60 (1.58-4.28)	
>50-60	44	9 564	4.60 (3.42-6.18)	3.61 (2.11-6.18)	
>60	22	3 601	6.11 (4.02-9.28)	4.79 (2.59-8.85)	
Current CD4 (cells/µl)					
<200	17	5 196	3.27 (2.03-5.26)	1	0.9
200-349	37	13 173	2.81 (2.04-3.88)	0.86 (0.48-1.52)	
350-499	69	22 420	3.08 (2.43-3.90)	0.94 (0.55-1.60)	
≥500	143	47 922	2.98 (2.53-3.52)	0.91 (0.55-1.51)	
Missing	0	760	-	-	
Nadir CD4 cell count (cells/µl)		(760 missing)			
<200	88	22 212	3.96 (3.21-4.88)	1	0.006
200-349	96	31 185	3.08 (2.52-3.76)	0.78 (0.58-1.04)	
350-499	47	19 228	2.44 (1.84-3.25)	0.62 (0.43-0.88)	
≥500	35	16 086	2.18 (1.56-3.03)	0.55 (0.37-0.81)	
Current HIV viral load (copies/mL)					
<1000	189	50 669	3.73 (3.23-4.30)	1	0.007
1000-9999	25	11 954	2.09 (1.41-3.09)	0.56 (0.37-0.85)	
10000-99999	44	14 820	2.97 (2.21-3.99)	0.79 (0.57-1.10)	
≥100000	8	4 738	1.69 (0.84-3.38)	0.45 (0.22-0.91)	
Missing	0	7 290	-	-	
Current ART[2]					
Not on ART	39	26 191	1.49 (1.09-2.04)	1	<0.0001
TDF[3], no PI[4]	38	12 108	3.14 (2.28-4.31)	2.11 (1.35-3.29)	
PI, no TDF	74	19 216	3.85 (3.06-4.84)	2.59 (1.75-3.81)	
PI & TDF	48	9 387	5.11 (3.85-6.79)	3.43 (2.25-5.24)	
Other ART (not PI/TDF)	67	22 568	2.97 (2.34-3.77)	1.99 (1.34-2.96)	
Duration of immune suppression (≤200cells/µL)					
CD4 always >200	178	66 499	2.65 (2.28-3.07)	1	0.03
>0-6 months	35	9 904	3.53 (2.54-4.92)	1.26 (0.88-1.82)	
>6-12 months	15	3 709	4.04 (2.44-6.71)	1.48 (0.88-2.51)	
>12 months	38	8 599	4.42 (3.22-6.07)	1.62 (1.14-2.30)	
Missing	0	760	-	-	

[1] Test for heterogeneity with missing excluded [2] ART antiretroviral therapy [3]TDF Tenofovir Disoproxil Fumarate
[4]Protease Inhibitor *PYFU-person-years of follow-up

Table 6.3 (Page 2 of 2): The association between time-varying factors and fracture incidence in PLHIV: Rates and univariable analysis

Time-updated factor	Event No	PYFU*	Rate/1000 PYFU (95%CI)	IRR (95%CI)	p[1]
Duration of immune suppression (≤100cells/μL)					
CD4 always >100	208	75 324	2.76 (2.41-3.16)	1	0.04
>0-6 months	5	1 839	2.72 (1.13-6.53)	0.98 (0.41-2.39)	
>6-12 months	15	3 709	3.71 (2.44-6.71)	1.46 (0.87-2.47)	
>12 months	38	8 599	4.42 (3.22-6.07)	1.60 (1.13-2.26)	
Missing	0	760	-	-	
Duration of immune suppression (≤50cells/μL)					
CD4 always >50	243	84 402	2.88 (2.54-3.26)	1	0.1
>0-6 months	11	2 562	4.29 (2.38-7.75)	1.49 (0.82-2.73)	
>6-12 months	12	2 507	4.79 (2.72-8.43)	1.66 (0.93-2.97)	
Missing	0	760	-	-	
Prior AIDS					
No	226	81 858	2.76 (2.42-3.15)	1	<0.0001
Yes	40	7 613	5.25 (3.85-7.16)	1.90 (1.36-2.66)	
HCV-seropositivity					
No	226	83 300	2.71 (2.38-3.09)	1	<0.0001
Yes	40	6 171	6.48 (4.75-8.84)	2.39 (1.71-3.34)	
Calendar period					
≤1996	8	6 160	1.30 (0.65-2.60)	1	0.009
>1996-2000	21	10 267	2.05 (1.33-3.14)	1.58 (0.70-3.56)	
>2000-2010	172	55 343	3.11 (2.68-3.61)	2.39 (1.18-4.86)	
>2010	65	17 700	3.67 (2.88-4.68)	2.83 (1.36-5.89)	

[1] Test for heterogeneity with missing excluded *PYFU-person-years of follow-up

Table 6.4 shows IRRs for the univariable analysis of continuous covariates. There was evidence of an association between each covariate and risk of fracture apart from time-updated CD4 cell count. For each additional 10 years of HIV infection fracture rates increased by more than 60%.There was also evidence that fracture rates increased with increasing current age (~40% per additional decade) and age at seroconversion (~25% per additional decade). There was an inverse association between fracture rates and $\log_{10}$ HIV viral load. Although there was evidence of increasing fracture rates with increasing calendar year, the association was found not to be linear as it plateaued in later calendar years. A categorical variable was therefore used for calendar period in the multivariable analysis.

Both age at seroconversion and current age were strongly positively correlated (Pearson's correlation coefficient (r_P)= 0.9). Age at seroconversion was weakly negatively correlated with duration of HIV infection (r_P =-0.1). Current age was moderately positively correlated with duration of HIV infection (r_P <0.4). Current age was included in preference to age at seroconversion as it was more strongly associated with fracture incidence, is a well-known predictor of fracture and the two variables could not both be included (see Section 4.4.2.3).

Table 6.4: The association between continuous covariates and fracture incidence in HIV-positive individuals: univariable analysis*

	IRR (95%CI)	p
Duration of HIV infection~	1.62 (1.30-2.02)	<0.0001
Seroconversion age~	1.24 (1.11-1.39)	<0.0001
Current age~	1.36 (1.22-1.51)	<0.0001
Nadir CD4 cell count (per 100 cell/µL increase)	0.89 (0.83-0.95)	0.001
Current Log HIV viral load (per log_{10} copy/mL increase)	0.85 (0.77-0.93)	<0.0001
Current CD4 cell count (per 100 cells/µL increase)	1.02 (0.97-1.06)	0.5
Duration of immune-suppression¥ (≤200 cells/mL)	1.09 (1.02-1.18)	0.01
Duration of immune-suppression¥ (≤100 cells/mL)	1.10 (1.02-1.18)	0.01
Duration of immune-suppression¥ (≤50 cells/mL)	1.23 (1.04-1.44)	0.01
Calendar year¥	1.05 (1.02-1.08)	<0.0001

* All variables were time-updated apart from age at seroconversion ~per additional 10 years ¥per additional year

6.5 Multivariable analysis

Table 6.5 shows the basic multivariable model A0 which included demographic factors (age, sex and ethnicity) and all available potential confounders determined *a priori* (ART, HCV-seropositivity, cohort and calendar period) (see section 2.3.4). All variables were significant at $p<0.05$ apart from sex and calendar period. A potential confounder not included in model A0 was mode of infection. Although, it was significant at $p<0.05$ in the univariable analysis both as nominal (MSM, PWID, heterosexual, other) and binary variable (PWID/non-PWID), mode was strongly associated with HCV-seropositivity (χ^2, $p<0.0001$ for both binary and categorical). I included -HCV-seropositivity as the two variables were not statistically distinguishable. HCV is a known risk factor for both fractures and their precursor osteoporosis. [291].

Table 6.5: The association between demographic factors/potential confounders and fracture incidence in HIV-positive individuals: Basic multivariable model A0

Variable	IRR (95%CI)	p
Current age~	1.25 (1.11-1.40)	<0.0001
Sex		
Male	1	-
Female	0.84 (0.62-1.14)	0.3
Cohort		
FRENCH[1]	1	<0.0001
AHIVCOS[2]	6.85 (4.92-9.54)	
NOR[3]	2.53 (1.19-5.35)	
SEROCO	2.13 (1.33-3.39)	
Apparent ethnicity‡		
Black/Other	1	-
White	2.03 (1.10-3.76)	0.02
ART[4] group		
Off ART	1	<0.0001
TDF[5]	1.84 (1.16-2.91)	
PI[6]	2.23 (1.49-3.32)	
PI & TDF	2.96 (1.91-4.57)	
Other ART	1.88 (1.26-2.81)	
HCV- seropositivity		
No	1	-
Yes	1.69 (1.19-2.40)	0.003
Calendar period		
≤1996	1	0.2
>1996-2000	1.39 (0.60-3.22)	
>2000-2010	1.96 (0.92-4.17)	
>2010	1.67 (0.74-3.74)	

[1]French cohort comprised FHDH and PRIMO combined (this merger of categories was appropriate as it improved the existing AIC score of the model by more than two points) [2]AHIVCOS-Austrian HIV cohort. [3]NOR-Oslo and Ulleval hospital cohort. [4]ART antiretroviral therapy [5]TDF Tenofovir disoproxil fumarate [6]PI protease Inhibitor ~per additional 10 years ‡based on region of origin

Interaction was then assessed between each variable included in model A0 and current age, sex and ethnicity. The interaction term p values are shown in Table 6.6. There was evidence for an interaction between current age and sex (p=0.04).

Table 6.6: The assessment of interaction between sex, age and ethnicity and other variables in the basic multivariable model A0
(examining the association between demographic factors/potential confounders and fracture incidence in PLHIV)

Variable	Interaction term p values*		
	Current age	Sex	Apparent Ethnicity
Sex	**0.04**	x	x
Cohort	0.6	0.3	0.8
Ethnic/region	0.8	0.7	x
ART	0.7	0.9	0.5
HCV-seropositivity	0.3	0.3	a
Calendar period	0.3	0.5	a

*Interaction terms were added individually to model A0 and tested for significance. X-not applicable a-could not be calculated as no events occurred in some categories

Table 6.7 shows the IRRs for current age stratified by sex. The IRR for females was ≥20% larger than that for males. There was no evidence of simultaneous confounding as the value of the combined IRR was between the values for males and females. If male and female IRRs had both been larger or smaller than the combined value this would have indicated simultaneous confounding. If both confounding and interaction had been operant it makes appropriate adjustment problematic. An interaction term between current age and sex was added to model A0.

Table 6.7: The assessment of interaction between sex and current age: Fitting the basic multivariable model A0 separately by sex

	Male[1] IRR (95%CI) p value	Female[1] IRR (95%CI) p value	Combined[2] IRR (95%CI) p value	p value for interaction
Current age~	1.16 (1.02-1.33) p=0.03	1.58 (1.25-1.92) p<0.0001	1.25 (1.11-1.39) p<0.0001	0.04

[1]Adjusted for cohort, ethnicity/region, ART and HCV-seropositivity [2]Adjusted for cohort, apparent ethnicity ART, HCV-seropositivity and sex ~per additional 10 years

In Table 6.8 all factors in multivariable model A were significant at p<0.05 apart from calendar period which was of borderline significance (p=0.1).

Table 6.8: The association between demographic factors/potential confounders and fracture incidence in HIV-positive individuals: Multivariable model A

Variable	IRR (95%CI)	p
Current age~	1.17 (1.02-1.33)	0.02
Sex		
Male	1	-
Female	0.26 (0.08-0.84)	0.02
Interaction term between age and sex (for females per 10 year increase in age)	1.31 (1.02-1.70)	0.04
Cohort		
FRENCH[1]	1	<0.0001
AHIVCOS[2]	6.83 (4.91-9.51)	
NOR[3]	2.57 (1.21-5.44)	
SEROCO	2.11 (1.32-3.37)	
Apparent ethnicity		
Black/Other	1	-
White	1.91 (1.03-3.54)	0.04
ART group		
Off ART	1	<0.0001
TDF	1.85 (1.24-3.05)	
PI	2.25 (1.51-3.35)	
PI & TDF	2.98 (1.93-4.61)	
Other ART	1.89 (1.26-2.83)	
HCV- seropositivity		
No	1	
Yes	1.71 (1.21-2.43)	0.002
Calendar period		
<1996	1	0.1
>1996-2000	1.38 (0.60-3.22)	
>2000-2010	1.96 (0.92-4.17)	
>2010	1.67 (0.74-3.74)	

[1]French cohort comprises FHDH and PRIMO combined. [2]AHIVCOS- Austrian HIV cohort [3]NOR-Oslo and Ulleval hospital cohort ~per additional 10 years

Table 6.9 shows IRRs for each covariate of interest added separately to model A and compares the values to the univariable IRRs. Duration of infection (model B) was no longer statistically significantly associated with fractures and the point estimate for the IRR after adjustment was close to one. Subsequent pair-wise analysis of the factors in model A determined that the attenuation of the association with duration of HIV infection was being driven by the inclusion of current age in the model and that the attenuation for nadir CD4 was being driven by the inclusion of the ART variable. When nadir CD4 cell count (model C), log HIV viral load (model D) or any measure of duration of immune-suppression (models F-H) were added to model A, the estimated IRR moved closer to 1 and the statistical significance of each association was attenuated. The only covariate of interest to retain statistical significance when added to model A was prior AIDS ($p=0.02$), which was associated with a 1.5-fold increased risk of fractures. Duration of immune suppression ≤50 cells/µL was associated with a 15% increase in fracture rates, but its p value of 0.12 was above the threshold for inclusion in the final model ($p<0.1$).

Table 6.9: The association between duration/markers of HIV infection and fracture incidence in HIV-positive individuals: IRRs for each covariate of interest added separately to multivariable model A (Model B-I) with univariable results shown for comparison

Model*	Covariate	Multivariable IRR (95%CI)*	p	Univariable IRR (95%CI)	p
B	**Duration of HIV Infection~**	1.03 (0.80-1.34)	0.8	1.62 (1.30-2.02)	<0.0001
C	**Current CD4 cell count (per 100 cell/µL increase)**	1.02 (0.97-1.06)	0.5	1.02 (0.97-1.06)	0.5
D	**Log viral load (per log copies/mL increase)**	0.95 (0.85-1.05)	0.3	0.85 (0.77-0.93)	<0.0001
E	**Nadir CD4 cell count (per 100 cell increase)**	0.99 (0.92-1.06)	0.8	0.89 (0.83-0.95)	0.001
F	**Years of immune suppression¥ (≤200cells/µL)**	1.01 (0.93-1.10)	0.8	1.09 (1.02-1.18)	0.01
G	**Years of immune suppression¥ (≤100cells/µL)**	1.01 (0.93-1.10)	0.8	1.10 (1.02-1.18)	0.01
H	**Years of immune suppression¥ (≤50cells/µL)**	1.15 (0.96-1.36)	0.1	1.23 (1.04-1.44)	0.01
I	**Prior AIDS (yes vs no)**	1.51 (1.07-2.12)	0.02	1.90 (1.36-2.66)	<0.0001

*Model A comprised: current age, sex, cohort, apparent ethnicity, ART group, HCV-seropositivity, calendar period with an interaction term between age and sex ¥ per additional year ~per additional 10 years

Duration of infection and prior AIDS were retained and added to model A because duration of infection was the primary covariate of interest and prior AIDS was a significant predictor at p<0.1 (Table 6.10).

Table 6.10: The association between duration/markers of HIV infection and fracture incidence in HIV-positive individuals after adjusting for demographic factors/potential confounders: The final multivariable model (model J)

Variable	IRR (95%CI)	p
Duration of HIV infection~	0.98 (0.76-1.28)	0.9
Current age~	1.16 (1.01-1.33)	0.03
Sex		
Male	1	0.03
Female	0.26 (0.08-0.85)	
Interaction term between age and sex (for females per additional 10 years)	1.31 (1.01-1.70)	0.04
Cohort		
FRENCH[1]	1	<0.0001
AHIVCOS[2]	6.81 (4.86-9.34)	
NOR[3]	2.53 (1.13-4.71)	
SEROCO	2.08 (1.14-2.80)	
Apparent ethnicity		
Black/other	1	0.04
White	1.93 (1.04-3.58)	
ART group		
Off ART	1	0.0001
TDF	1.80 (1.13-2.86)	
PI	2.13 (1.41-3.21)	
PI & TDF	2.88 (1.85-4.48)	
Other ART	1.83 (1.21-2.76)	
Prior AIDS		
No	1	0.02
Yes	1.51 (1.07-2.14)	
HCV-seropositivity		
No	1	0.003
Yes	1.69 (1.19-2.42)	
Calendar period		
<1996	1	0.2
1996-2000	1.39 (0.60-3.24)	
2000-2010	1.99 (0.93-4.26)	
>2010	1.72 (0.76-3.89)	

[1]French cohort comprises FHDH and PRIMO combined. [2]AHIVCOS- Austrian HIV cohort [3]Norway-Oslo and Ulleval hospital cohort ~per additional 10 years

All previously excluded covariates were then added back into the model one at a time; none met the criteria for re-inclusion. Previously unassessed interactions were then evaluated, but no further significant interactions ($p<0.1$) were found.

In the final model (Table 6.10) duration of infection was retained as it was the primary exposure of interest, although there was no evidence of a statistically significant association with fractures and the point estimate for the IRR was very close to one. Prior AIDS led to a ~50% increase in fracture rates and was the only covariate of interest found to be statistically associated with fractures after adjustment.

6.6 Sensitivity analyses

A number of sensitivity analyses were undertaken to explore the robustness of assumptions.

Results of sensitivity analysis 1a which investigated the effects of censoring follow-up after the first gap in CD4 measurements of at least a year (Table 6.11) were very similar to those found in the main analysis (Table 6.9) and none of the covariates derived from CD4 cell counts were eligible for inclusion in the final model (although duration of immune suppression ≤100cells/μL was close to the cut-off of p=0.1).

Table 6.11: Sensitivity analysis 1a: Each CD4 related covariate added separately (model C/E-H) to model A after censoring follow-up after the first gap in CD4 measurements of ≥1 year

Sensitivity model	Covariate	Main analysis IRR (95%CI)*	p	Sensitivity model 1_a IRR (95%CI)*	p
C_{1a}	**Current CD4 cell count Per 100 (cell/μL increase)**	1.02 (0.97-1.06)	0.5	1.01 (0.96-1.06)	0.7
E_{1a}	**Nadir CD4 cell count (per 100 cell increase)**	0.99 (0.92-1.06)	0.8	1.02 (0.97-1.08)	0.5
F_{1a}	**Years of immune suppression¥ (≤200cells/μL)**	1.01 (0.93-1.10)	0.8	1.04 (0.93-1.16)	0.5
G_{1a}	**Years of immune suppression¥ (≤100cells/μL)**	1.01 (0.93-1.10)	0.8	1.14 (0.96-1.35)	0.1
H_{1a}	**Years of immune suppression¥ (≤50cells/μL)**	1.15 (0.96-1.36)	0.1	1.14 (0.87-1.48)	0.4

*Each model also included: current age, sex, cohort, apparent ethnicity, ART group, HCV-seropositivity , calendar period and an interaction term between age and sex ¥per additional year

Results of sensitivity analysis 1b which built on 1a, censoring follow-up after each gap in CD4 measurements of at least a year but allowing re-entry if there were subsequent CD4 measurements were similar to the main analysis for models C, E and F, but models G and H differed from the main analysis (Table 6.12). The estimated IRRs were higher in model G_{1b} and H_{1b} compared to G and H and the associations between time spent with severe immune-suppression at both ≤100 cell/μL and ≤50 cell/μL and fracture incidence was of borderline significance (p=0.06 for both immunosuppression levels). Further investigation found that higher CD4 cell counts were more likely to be carried forward for extended periods (data not shown).

Table 6.12: Sensitivity analysis 1b: Each CD4 related covariate added separately (model C/E-H) to model A after censoring follow-up after the first gap in CD4 measurements of ≥1 year, but allowing re-entry to follow-up if subsequent CD4 measurements were recorded

Sensitivity model	Covariate	Main analysis IRR (95%CI)*	p	Sensitivity model 1_a IRR (95%CI)*	p
C_{1b}	**CD4 cell count (per 100 cell/μL increase)**	1.02 (0.97-1.06)	0.5	1.01 (0.97-1.06)	0.6
E_{1b}	**Nadir CD4 cell count (per 100 cell increase)**	0.99 (0.92-1.06)	0.8	0.99 (0.92-1.06)	0.8
F_{1b}	**Years of immune suppression¥ (≤200cells/μL)**	1.01 (0.93-1.10)	0.8	1.04 (0.95-1.13)	0.4
G_{1b}	**Years of immune suppression¥ (≤100cells/μL)**	1.01 (0.93-1.10)	0.8	1.13 (0.99-1.30)	0.06
H_{1b}	**Years of immune suppression¥ (≤50cells/μL)**	1.15 (0.96-1.36)	0.1	1.20 (0.99-1.45)	0.06

*Each model also included: current age, sex, cohort, apparent ethnicity, ART group, HCV-seropositivity , calendar period and an interaction term between age and sex ¥ per additional year

Sensitivity analyses assessing the impact of carrying HIV viral load forward for more than a year are shown in Table 6.13 and Table 6.14. These results were very similar to those found in the main analysis and log HIV viral load remained ineligible for inclusion in the final model due to lack of evidence for a statistically significant association with fractures.

Table 6.13: Sensitivity analysis 2a: HIV viral load added to model A after censoring follow-up after the first gap in HIV viral load measurements of ≥1 year

Model*	Covariate	Multivariable IRR (95%CI)*	p	Sensitivity model 2_a IRR (95%CI)*	p
D_{2a}	Log HIV viral load (per log_{10} increase)	0.95 (0.85-1.05)	0.3	0.93 (0.82-1.05)	0.2

*The model also included: current age, sex, cohort, apparent ethnicity, ART group, HCV-seropositivity, calendar period and an interaction term between age and sex

Table 6.14: Sensitivity analysis 2b: HIV viral load added to model A after censoring follow-up after the first gap in HIV viral load measurements of ≥1 year, but allowing re-entry to follow-up if subsequent HIV viral load measurements were recorded

Model*	Covariate	Multivariable IRR (95%CI)*	p	Sensitivity model 2_b IRR (95%CI)*	p
D_{2b}	Log HIV viral load (per log_{10} increase)	0.95 (0.85-1.05)	0.3	0.83 (0.82-1.04)	0.2

*The model also included: current age, sex, cohort, apparent ethnicity, ART group, HCV-seropositivity, calendar period and an interaction term between age and sex

Sensitivity analysis 3 restricted the analysis to ≥01/01/2005 when cohort investigators were more likely to be aware of the importance of collecting data on fractures in PLHIV. There were 177 fractures and 11 800 individuals included in the analysis; this was 89 fewer fractures and 1 443 fewer individuals than the main analysis. Total PYFU was approximately 57% of that of the main analysis (50 854 years compared to 89 470) and the crude first fracture rate was 3.5 (3.0-3.6) fractures per 1000 PYFU (vs. 3.0 (2.6-3.4) in the main analysis). In Table 6.15 the results from sensitivity analysis 3 are compared to those from the main analysis for models B-I.

Table 6.15: Sensitivity analysis 3: Each covariate of interest added separately to multivariable model A (models B3-I3) after restricting follow-up to ≥01/01/2005 (IRRs from the main analysis also included)

Sensitivity model	Covariate	Main analysis IRR (95%CI)*	p	Sensitivity model 3 IRR (95%CI)*	p
B_3	**Duration of HIV Infection~**	1.03 (0.80-1.34)	0.8	1.01 (0.75-1.35)	0.9
C_3	**CD4 cell count (per 100 cell/µL increase)**	1.02 (0.97-1.06)	0.5	1.02 (0.97-1.08)	0.4
D_3	**Log viral load (per log_{10} copies/mL increase)**	0.95 (0.85-1.05)	0.3	0.98 (0.85-1.12)	0.8
E_3	**Nadir CD4 cell count (per 100 cell/µL increase)**	0.99 (0.92-1.06)	0.8	0.99 (0.90-1.09)	0.9
F_3	**Duration of immune suppression¥ (≤200cells/µL)**	1.01 (0.93-1.10)	0.8	1.00 (0.91-1.10)	1.0
G_3	**Duration of immune suppression¥ (≤100cells/µL)**	1.01 (0.93-1.10)	0.8	1.00 (0.91-1.11)	0.9
H_3	**Duration of immune suppression¥ (≤50cells/µL)**	1.15 (0.96-1.36)	0.1	1.16 (0.96-1.42)	0.1
I_3	**Prior AIDS (yes vs no)**	1.51 (1.07-2.12)	0.02	1.51 (1.00-2.29)	0.05

*Adjusted for: current age, sex, cohort, apparent ethnicity, ART group, HCV-seropositivity, calendar period with an interaction term between age and sex.¥ per additional year ~per additional 10 years

When each covariate of interest was added to model A in sensitivity analysis 3 (Table 6.15) the results were very similar to those of the main analysis, but with slightly wider confidence intervals due to the smaller sample size.

When model J (the main analysis final model) was fitted to the amended data the IRRs for duration of HIV infection and prior AIDS (the only two covariates of interest included) were almost identical to those of the main analysis, IRR=0.96 (0.71-1.29, p=0.8) and IRR=1.53 (1.00-2.33, p=0.02) respectively.

In sensitivity analysis 4, individuals were not permitted to enter follow-up until the day after the first event of that type was recorded by their cohort. Four fractures and 44 individuals were excluded from the analysis as a result of this change. It also reduced total

PYFU from 89 470 years in the main analysis to 82 598 years. The crude first fracture rate was 3.2 (2.8-3.6) fractures per 1000 PYFU compared to 3.0 (2.6-3.4) in the main analysis. In Table 6.16 the results from this sensitivity analysis are compared to those from the main analysis for models B_x-I_x.

Table 6.16: Sensitivity analysis 4: Each covariate of interest added separately to multivariable model A (models B_4-I_4) after amending the start of follow-up for each individual so they could not enter before the day after the first recorded event for their cohort (IRRs from the main analysis also included)

Sensitivity model	Covariate	Main analysis IRR (95%CI)*	p	Sensitivity model 4 IRR (95%CI)*	p
B_4	**Duration of HIV Infection~**	1.03 (0.80-1.34)	0.8	1.04 (0.80-1.35)	0.8
C_4	**Current CD4 cell count (per 100 cell/µL increase)**	1.02 (0.97-1.06)	0.5	1.02 (0.97-1.06)	0.4
D_4	**Log viral load (per log_{10} copy/mL increase)**	0.95 (0.85-1.05)	0.3	0.94 (0.84-1.04)	0.2
E_4	**Nadir CD4 cell count (per 100 cell/µL increase)**	0.99 (0.92-1.06)	0.8	0.99 (0.92-1.06)	0.8
F_4	**Duration of immune suppression¥ (≤200cells/µL)**	1.01 (0.93-1.10)	0.8	1.01 (0.93-1.10)	0.8
G_4	**Duration of immune suppression¥ (≤100cells/µL)**	1.01 (0.93-1.10)	0.8	1.01 (0.93-1.10)	0.8
H_4	**Duration of immune suppression¥ (≤50cells/µL)**	1.15 (0.96-1.36)	0.1	1.15 (0.97-1.37)	0.1
I_4	**Prior AIDS (yes vs no)**	1.51 (1.07-2.12)	0.02	1.52 (1.08-2.15)	0.02

*Adjusted for: current age, sex, cohort, apparent ethnicity, ART group, HCV-seropositivity, calendar period with an interaction term between age and sex. # The start of follow-up for each individual either began the day after the first event of that type was recorded by their cohort or at enrolment whichever happened later. ¥ per additional year ~per additional 10 years

When each covariate of interest was added to model A in sensitivity analysis 4 (Table 6.16) the results were very similar to the values in the main analysis with respect to both the estimated IRR and the p-value.

When model J (the final in the main analysis) was refitted to the data amended to exclude the day of the first event the IRR for duration of HIV infection was almost unchanged from the main analysis, IRR=0.99 (0.76-1.29, p=0.9). Similarly the effect of prior AIDS, IRR=1.53 (1.08-2.16, p=0.02) was also very close to that of the main analysis.

The final two sensitivity analyses (sensitivity analyses 5 and 6) examined the possibility of non-linear associations between covariates and fractures using mfp.

The first sensitivity analysis (analysis 5) explored potential non-linear associations in model J (the final model from the main analysis).

Table 6.17: Sensitivity analysis 5: Assessing the effect of fractional transformations (mfp) of variables include in the final multivariable analysis (model J) on the association between duration/markers of HIV infection and fractures*

Variable	Analysis	vs.	Deviance	Deviance difference	p	Powers	vs.
Duration of Infection~	Linear	FP2	5266	2.3	0.5	1	-2 -2
	Final		5266			1	
Current age~	Linear	FP2	5266	2.6	0.4	1	-2 -2
	Final		5266			1	

*The model also included: sex, cohort, apparent ethnicity, ART group, HCV-seropositivity, calendar period ~per additional 10 years

Table 6.17 compares the inclusion of the best fitting FP2 transformation (see Table 4.4, Methods for all transformations considered) to the variable in its linear form for the two continuous covariates included in the final multivariable analysis (model J) of the main analysis. For both duration of HIV infection and current age the best-fitting FP2 transformation (of the 36 tested) was found to be a $1/x^2$ transformation for both p^1 and p^2 (where FP2 takes the form $\beta_1 X^{p1}+\beta_2 X^{p2}$, see methods chapter), but in neither case was there evidence that the FP2 transformation was a better fit than the linear term (p=0.5 and p=0.4 respectively).

In the second sensitivity analysis (analysis 6) all available variables (see Table 4.4, Methods) were initially included in the model with the following exceptions: duration of immune-suppression ≤100 and ≤50 cells/μL were omitted as collinearity prevented >1 duration of immune-suppression variable from being included simultaneously and ≤200 cells/μL was most clinically relevant (since the introduction of cART very severe immunosuppression has become less common); age at seroconversion was omitted due to its collinearity with current age and mode of infection was omitted due to its strong association with HCV-seropositivity.

Table 6.18: Sensitivity analysis 6: The association between duration/markers of HIV infection/other factors and fracture incidence in HIV-positive individuals: mfp combined with backwards elimination (final cycle) [#]

Variable	Model	vs.	Deviance	Deviance difference	p	Powers	vs.
Duration of HIV Infection~	Linear	FP2	5245	2.0	0.6	1	-1 -1
	Final		5245			1	
Current age~	Null	FP2	5257	14.9	0.005	-	-2 -2
	Linear		5245	2.7	0.4	1	
	Final		5245			1	
Nadir CD4 cell count	Null	FP2	5245	1.1	0.9	-	0 0.5
(cells/µL*)	Final		5245				
Current CD4 cell count	Null	FP2	5245	2.6	0.6	-	2 2
(cells/µL*)	Final		5245			-	
Duration of immunosuppression¥ (≤200 cells/µL*)	Null	FP2	5245	0.5	1.0	-	-2 -1
	Final		5245				
Log_{10} HIV viral load	Null	FP2	5245	1.1	0.9		-2 -2
(per log_{10}copy/mL increase)	Final		5245				

[#]The following binary and categorical variables were also assessed for inclusion: prior AIDS, prior HCV, calendar period, cohort; apparent ethnicity, mode, sex and ART group.*All units are those for the linear form of the variable ~per additional 10 years ¥per additional year

Table 6.18 shows the results of combining mfp for continuous variables with backwards-stepwise selection for all variables. Duration of HIV infection was forced into the model and all other exposures of interest were eliminated if p≥0.1 during the stepwise process. The associations between all continuous variables and the outcome were found to be best described by the variable in its linear form. No covariates of interest were included in the final analysis as they were not statistically significant at p<0.1. The exception to this was duration of HIV infection, which as the primary covariate of interest was retained regardless of its statistical association.

Table 6.19: Sensitivity analysis 6: The associations between duration/markers of HIV infection and fracture incidence in HIV-positive individuals: mfp combined with backwards elimination (final model)

Variable	IRR (95%CI)	p
Duration of Infection~	0.98 (0.76-1.26)	0.9
Current age~	1.24 (1.10-1.40)	<0.0001
Apparent ethnicity (white vs black/other)	2.17 (1.18-3.98)	0.01
Cohort		
FRENCH[1]		
AHIVCOS[2]	6.59 (4.75-9.15)	<0.0001
NOR[3]	2.25 (1.10-4.59)	0.03
SEROCCO	1.70 (1.09-2.67)	0.02
HCV-seropositivity (yes vs no)	1.65 (1.162.35)	0.005
Prior AIDS (yes vs. no)	1.52 (1.07-2.15)	0.02
ART[4] group		
Off ART	1	-
TDF[5]	1.63 (1.03-2.58)	0.04
PI[6]	1.92 (1.28-2.88)	0.002
PI and TDF	2.62 (1.70-4.06)	<0.0001
Other ART	1.67 (1.11-2.51)	0.02

[1]French cohort comprises FHDH and PRIMO combined. [2]AHIVCOS- Austrian HIV cohort [3]NOR-Oslo and Ulleval hospital cohort excluded [4]ART antiretroviral therapy [5]TDF Tenofovir disoproxil fumarate [6]PI protease Inhibitor
~per additional 10 years

Table 6.19 shows the results of sensitivity analysis 6. There was no evidence that any covariates of interest (CD4, viral load, duration of infection/immune suppression) except prior AIDS were statistically significantly ($p<0.05$) associated with fractures after adjustment. All other covariates of interest included initially in the backwards-stepwise model were eliminated. These findings were in agreement with the main analysis. The main difference between sensitivity analysis 6 and the findings of the main analysis was that calendar period and sex were excluded during backwards elimination in sensitivity analysis 6.

6.7 Summary of my findings

The results are summarised in Table 6.20. With the exception of current CD4 cell count, there was evidence of an association between all covariates of interest, including duration of HIV infection and fracture incidence in the univariable analysis. For each additional 10 years of HIV infection duration, the unadjusted first fracture incidence increased by approximately 60%. Once current age was adjusted for, however, there was no evidence of an association between duration of HIV infection and fractures. Prior AIDS was the only covariate of interest for which statistical evidence of an association with fractures was retained after adjustment. Individuals with a prior AIDS diagnosis experienced a 50% increase in fracture incidence compared to those without one after accounting for other factors.

All findings were robust to all sensitivity analyses with the exception of duration of immune suppression (≤100 and ≤50 cells/μL). These two measures of immune suppression were not statistically significantly associated with fracture incidence after adjustment in the main analysis, but were of borderline significance in sensitivity analysis 1b. In this analysis follow-up was censored after each gap of more than a year between CD4 measures (with re-entry at subsequent CD4). Further investigation found that higher CD4 cell counts were more likely to be carried forward over long periods of time in the main analysis than lower CD4 cell counts. This would make sense clinically as doctors would probably re-test those with known low CD4 cell counts more frequently. In the sensitivity analysis, these periods (included in the main analysis) where high CD4 cell counts were carried forward for a long time and events occurred were removed. This then increased the apparent association between duration of immune suppression and fractures.

Overall, there was little evidence from my analysis that any HIV-related variable was associated with fracture incidence after adjustment (except prior AIDS), and no evidence of such an association for my primary exposure of interest, duration of HIV infection.

Table 6:20: Summary of fracture analysis findings: The association between markers of HIV infection and fracture incidence before and after adjustment and during sensitivity analysis

Exposure of Interest*	Univariable	Adjusted~	Final Multivariable#	Sensitivity Analysis	Comments
Duration of HIV infection (per additional 10 years)	**1.62 (1.30-2.02) p<0.0001**	1.03 (0.80-1.34) p=0.8	0.98 (0.76-1.28) p=0.9	All results similar size/direction to main multivariable analysis	Univariable association attenuated after adjustment for current age
Nadir CD4 cell count (per 100 cell increase)	**0.89 (0.83-0.95) p=0.001**	0.99 (0.92-1.06) p=0.8	-	All results similar size/direction to main multivariable analysis	Univariable association attenuated after adjustment for time-updated ART (TDF/PI/TDF & PI/ Other ART)
Current CD4 cell count (per 100 cell increase)	1.02 (0.97-1.06) p=0.5	1.02 (0.97-1.06) p=0.5	-	All similar size/direction to main multivariable analysis	
Time spent ≤200 cells/µL (per additional year)	**1.09 (1.02-1.18) p=0.01**	1.01 (0.93-1.10) 1.02 p=0.8	-	All results similar size/direction to main multivariable analysis	Univariable association attenuated after adjustment for current ART (both TDF/ PI)
Time spent ≤100 cells/µL (per additional year)	**1.10 (1.02-1.18) p=0.01**	1.01 (0.93-1.10) 1.02 p=0.8	-	**SA 1b: 1.13 (0.99-1.30) p=0.06**	SA 1b (borderline significance): Follow-up was censored after any gap of >1 year between CD4 measurements, with analysis re-entry if subsequent CD4 measurements were taken.
Time spent ≤50 cells/µL (per additional year)	**1.23 (1.04-1.44) p=0.01**	1.15 (0.96-1.36) p=0.1	-	**SA 1b: 1.20 (0.99-1.45) p=0.06**	
Current HIV viral Load (per log_{10} Cps/mL increase)	**0.85 (0.77-0.93) p<0.0001**	0.95 (0.85-1.05) p=0.3	-	All results similar size/direction to main multivariable analysis	Univarable association attenuated after adjustment for current ART (both TDF/PI)
Prior AIDS diagnosis	**1.90 (1.36-2.66) p<0.0001**	**1.51 (1.07-2.12) p=0.02**	**1.51 (1.07-2.14) p=0.02**	All results similar size/direction to main multivariable analysis	Evidence for an independent association between prior AIDS and fracture incidence

*All variables were time-updated ~Each adjusted model (models B-I from Table 6.9) was adjusted for the following potential confounders: current age, sex (with interaction term between sex/age), cohort, assumed ethnicity, ART, HCV and calendar period. # Results for the final multivariable model (J) which included duration of HIV infection and prior AIDS and was adjusted for all potential confounders previously included in the adjusted model

6.8 A comparison of my study and other studies

Table 6.21a and 6.21b compare study characteristics and analysis features of studies which examined the association between relevant exposures and fractures (including my own). Data from HOPS were included in two studies [182, 198], but no double reporting occurred.

6.8.1 Study design and population

All studies had a cohort design with the exception of a single matched nested case-control study undertaken in Australia [178]. Most studies analysed routine clinical data passively captured. Some US studies examined data from databases of healthcare insurance providers [132, 184]). The ALLRT, WIHS, MACS and SUN cohorts (as well as PRIMO and SEROCO in my study) included at least some active data capture [186, 187, 195, 196, 234]. All ALLRT patients were enrolled in AIDS Clinical Trial Group Trials (ACTG) [292]. The WIHS and SUN studies scheduled visits every six months and included a detailed history, physical examination and referral/testing as needed [198, 293, 294].

The majority of my patients were resident in France (>90%), the rest from Norway and Austria. All other studies were undertaken in the USA with the exception of one Swiss [185] and one Australian study [178].

6.8.2 Sampling (ascertainment)

Biased sampling can impact the external validity (generalisability) of findings [295]. The Danish Hospital Cohort study (DHCS) was at the lowest risk of sampling bias. It includes all HIV-positive patients receiving care in Denmark. Robust medical data linkage provides near complete health information [131, 296].

The US ALLRT study included individuals enrolled in AIDS Clinical Trial Group (ACTG) trials [187, 292]. Patients actively chose to participate and certain high-risk groups were excluded [297-299]. WIHS is a women-only cohort of HIV-positive and negative women (at high-risk of HIV) [186, 300]. MACS is restricted to MSM with or without HIV and only about 30% of those enrolled in MACS were included in the study [195]. In the HOPS and the Boston study white individuals were over-represented compared to PLHIV in the US [182, 197, 301]. In the HOPS/SUN study the majority had private medical insurance (56%), so were relatively wealthy. The VACS-VC study was restricted to male US veterans [184]. Numbers of MSM and PWID were not stated or adjusted for due to prohibition/stigma in the US military [302]. The SHCS population is representative of PLHIV in Switzerland [303]. The Australian case-control study sample [178] had a similar distribution of demographic characteristics to that of the Australian HIV-positive population [304, 305].

6.8.3 Loss to follow-up

The interval cohorts (without routine data access) were at higher risk of informative censoring than the clinical cohorts as ill patients are at increased risk of drop out [292]. If those who dropped were both exposed and had a fracture post drop out (which would not be included in the study), this could have pushed apparent study findings towards the null.

The DHCS has almost negligible loss [296, 306] with just 0.4% unaccounted for since cohort inception [131]. Amongst other studies loss to follow-up varied between 2% (WIHS and VACS-VC) and 10% (HOPS) annually. My study lost 4% each year. Annual loss to follow-up was not reported for the SUN/HOPS study [198] or the Boston Hospital study [197].

6.8.4 Survivor bias

The inclusion criteria for my study are likely to have produced some selection bias. It is difficult to assess the possible degree of survivor bias in other studies. The DHCS may have been at lower risk due to the near complete data capture on events from enrolment. Studies which only started capturing fracture data in later calendar years (e.g. SHCS and ALLRT [185, 187]) may be at increased risk.

6.8.5 Time-interval bias

Other studies only provided mean or median follow-up times. Average follow-up was <4 years for the SUN/HOPS, Boston, SHCS and HOPS studies [182, 185, 197, 198]. Short follow-up could push findings erroneous towards the null if there is a substantial time-lag between cause and effect.

6.8.6 Case definitions, misclassification and under-reporting

6.8.6.1 *Case definitions*

Table 6.21b includes a summary of outcome definitions and outcome identification methods across studies. Fractures are usually straightforward to diagnose if clinicians take a systematic history, clinical examination and imaging evaluation. X-ray errors/discrepancies occur in ~3-5% of cases [307]. Certain fragility fractures, such as those of the spine, are commonly missed [308].

The fracture case definition varied between studies, which could have contributed to the observed heterogeneity in findings. The type of fractures included varied. Studies either examined low-impact (fragility) fractures [132, 178, 185, 195] or included a combination of both low and high [182, 187, 196-198]. Fragility fractures were defined either by the level of traumatic impact [185] or by the fracture site [132, 178, 195]. I was not able to

distinguish between fragility, traumatic and pathological fractures in my study. There were five other studies who examined all fractures together [182, 187, 196-198]. The associations between HIV-related exposures and fracture type (high/low impact) may differ and could have contributed to a lack of agreement in findings between studies.

Whilst most studies included first (known) fracture only, the Boston Hospital study and WIHS included multiple events per person [186, 196, 197]. This might lead to bias as statistical methods usually assume events are independent. Prior fracture (especially fragility) is known to be an important risk-factor for future fracture [309].

6.8.6.2 *Misclassification/under-reporting*

The risk of misclassification also varied between studies (Table 6.21b). It can depend on the data source and the extent of data linkage. Endpoint review and validation have been shown to reduce misclassification for non-AIDS events [19, 285]. Three interval cohort studies (ALLRT, WIHS, MACS), relied on patient self-report [186, 187, 195, 196]. Patient recall of previous fracture has been found to be reasonably accurate (<10% false positives and negatives), but varies by fracture site [310, 311].

Data sources varied. For the WIHS and MACS analyses, laboratory test results, clinical findings and questionnaire responses collected at scheduled visits were recorded in a database [195, 196]. For the ALLRT study (also an interval cohort) information collected from the parent clinical trial was also available [187]. Most other clinical cohorts relied on electronic medical records from the clinic. Relevant ICD codes were used for the Boston (& problem list), VACS-VC, DHCS and Alfred Hospital studies [131, 132, 178, 184, 197]. The SHCS used detailed case report forms (CRFs). [185]. The HOPS/SUN and HOPS only study abstracted outcomes from EMR, but the paper did not state how [182, 198].

Under-reporting can lead to bias, if the likelihood of reporting differs by some combination of exposure and outcome status [312]. Under-reporting can be affected by the location of fracture [308]). A lack of linkage between HIV clinics and other medical facilities can lead to under-reporting. Due to effective linkage the DHCS was at very low risk [131]. Under-reporting appears likely in my study. VACS-VC had access to insurance information [132, 184, 313]. It was unclear how much of a potential problem under-reporting might be in most studies. The HOPS and ALLRT publications discussed under-reporting as a likely limitation [182, 187].

6.8.7 Confounding

The potential confounders adjusted for across studies were highly variable (Table 6.21b). For time-varying factors some studies adjusted for a time-fixed value (e.g. at index) and others undertook time-updated adjustment.

Factors associated with fractures are much more well-established than factors associated with my exposures of interest [314]. Important additional risk factors in PLHIV include HCV infection and exposure to ART (especially PIs and TDF) [315]. It HIV is causally associated with fractures then it is likely to increase fracture risk via loss of BMD [132, 316-319]. If this is the case it should not be adjusted for as it is on the causal pathway.

A number of studies (SUN/HOPS, Boston, ALLRT and Alfred Hospital) adjusted for low BMD or bisphosphate (low BMD treatment) [178, 187, 197, 198]. The inclusion of these variables in a multivariable model could incorrectly push findings toward the null [320]. The WIHS, Boston and SUN/HOPS studies [183, 196, 197] adjusted for prior fractures at index. If these occurred after HIV seroconversion then adjustment for them could attenuate valid associations.

6.8.8 Statistical Analysis

The VACS-VC and MACS studies undertook multiple imputation of missing covariates [132, 195]. All other studies performed a complete case analysis. Of these, only Boston (BMI), HOPS (BMI, ethnicity & ART) and my own (all) reported the amount of missing data for variables [182, 195].

Studies varied with respect to how factors were selected for inclusion in the final model. The majority of studies selected *a priori* [184, 185, 195]. WIHS, ALLRT used a backwards-stepwise selection process ($p<0.1$) [186, 187]. The method used by the Australian study was ambiguous [178]. Boston, HOPS/SUN and my study used a combination of *a priori* variable selection and selection based on evidence for relevant associations at the analysis stage [182, 197, 198].My study, MACS, VACS-VC and the SUN/HOPS study included time-updated covariates [132, 184, 195, 198]. Other studies were at risk of residual confounding for factors such as age. The SHCS and DHSCS were the only study to account for the competing risk of death [185]. No studies addressed informative censoring or time-varying confounding statistically.

6.8.9 Sensitivity analyses

WIHS, Boston and the Alfred Hospital studies did not report undertaking any sensitivity analyses [178, 186, 197]. I undertook stratification (age, sex, & ethnicity) as did DHCS (HCV, cART status and fracture type) and ALLRT (sex and cART status) and presented results for these separately [131, 187]. The SHCS included a competing risks analysis [185]. In a sensitivity analysis the VACS Index variable (a validated composite measure of frailty) was replaced by its component parts (including time-updated HIV viral load and CD4) [184]. This study also undertook a complete case analysis (multiple imputation was used in the main analysis) and findings were similar to the main results [184].The MACS study restricted the analysis to fragility fractures only [195].

Table 6.21a: Characteristics of studies examining the association between HIV-related factors & fractures

	My study	SUN/HOPS [183, 198]	Boston Hospitals [197]	SHCS [185]	WIHS [128] [186]	VACS-VC [132, 184]	ALLRT [187]	Alfred Hospital [178]	HOPS [182, 313]	MACS [195]	DHCS [131]
STUDY CHARACTERISTICS											
Study Type	Collaboration	Clinical cohort	Clinical cohort	Clinical cohort	Interval Cohort	Clinical cohort	Interval cohort [b]	Case-control [c]	Clinical cohort	Interval cohort	National cohort
Country/ Region	Mostly EU	USA	USA	Switzerland	USA	USA	USA	Australia	USA	USA	Denmark
Cohort type	Clinical and Interval [a]	Clinical (8 Sites)	Clinical (2 Sites)	Clinical (7 Sites)	Clinical (6 Sites)	Clinical (8 Sites)	Cohort of pooled RCTs	Clinical (1 Site)	Clinical (10 Sites)	Clinical (4 Sites)	Clinical (8 Sites)
Analysis design	Cohort	Cohort	Cohort	Cohort	Cohort	Cohort	Cohort	Case-control[c]	Cohort	Cohort	Cohort
FU period	1985- 2013	2004-2012	2001-2011	2008- 2010	2002-2013	1997-2009	2009-2011	1998-2009	2000-2008	2001-2015	1995-2009
DEMOGRAPHY											
Male	77%	83%	72%	70%	0%	100%	83%	89%	79%	100%	76%
Black	11%	21%	Not stated	8%[321]	56%	63%	29%	4%	33%	27%	14%
MSM	61%	69%	Not stated	43%	0%	Not stated	Not stated	Not stated	58%	100%	45%
PWID	4%	6%	Not stated	17%	3%	Not stated	10%	2%	14%	(1%)	11%
Annual LTFU	4%	Not stated	Not stated	Find	2%	2%	4%	n/a	10%	3%	0.4%
Data source (outcome)	MR	MR/ Interval	EMR	EMR/CRFs	Self-report	EMR/Survey/ Insurance	Self-report	MR [d]	MR/Self-report	Self-report	Linked EMR
Data source (other)	(E)MR/ registry	MR/ Interval*	EMR	EMR	Interval data	EMR/ Self-report	CRF/ Self-report	MR	MR	Self-report/ interval	MR/ Registry Migration
PYFU [IQR]≠	5.3 [2.5-10.1]	3.2 [1.7-6.5]	3.0 (SD, 2.5)	2.6†	10†	*6.0 [+/-3.9]*	5.0†	n/a	3.8 [1.5-7.4]	*13*†	6.5 [2.5-12.1]

ICD-International Classification of Diseases (E)MR-(Electronic) medical records RCT-randomised controlled trial SD-Standard deviation ≠Median PYFU (unless in italics which denotes mean)
*SUN only [a] PRIMO and SEROCO cohorts are interval cohorts [b] Patients seen every 16 weeks, previous enrolment in AIDS Clinical trial group RCTs [c] nested in a cohort [d] cases identified through the state HIV database & clinic records included †IQR/95%/SE CI not reported

Table 6.21b: Characteristic of studies examining the association between HIV-related factors & fractures

	My study	SUN/HOPS [183, 198]	Boston [197]	SHCS [185]	WIHS [128] [186, 196]	VACS-VC [132, 184]	ALLRT [187]	Alfred Hospital[178]	HOPS [182]	MACS [195]	DHCS[131]
OUTCOMES											
Means of ID	Recorded in (E)MR	Abstracted from MR [a]	ICD-9 in MR	CRF	Self-report	ICD-9 in MR	Self-report	ICD-9 in MR	MR/self-report	Self-report (ICD)	EMR (ICD-10)
Fracture type	All	All	All	All/ Fragility	All	Fragility	All	Fragility	All	Fragility	All/ Fragility
Validation/ Review	No/No	No/No	Yes /No	Yes	Yes	Yes	No/No	No/Yes	No/No	No/No	No/No
Events	First event only	First event only	Multiple events	First event only	Multiple events	First only	First event only	First event only	First event	First (aged >40 years)	First
Crude rate /1000 PYFU	3.0 (2.6-3.4)	2.1 (1.7-2.6)	22.2 (17.9–26.6)	1.64 (1.19-2.26)	21.9	-	4.0 (3.3-4.8)	5.3 (4.3-6.5)	-	12.8 (11.1-14.8)	21.0 (19.8-22.2)
POTENTIAL CONFOUNDERS											
Not captured	1-8, 12,14		4, 12,14	7-10,12,14	4,5,7,14	3, 4, 7	3,4,6,8,12	1,3,4,6,12	3,4,7,12,	3-5, 7, 12	2-8
Time-updated	0,10,11	11	11	0,1	-	0,5,10,14,	1,11	-		11	-
Time-fixed or excluded	9, 13	0,2,3,6,7,9, 10, 13	1-3,5-9,10,13	2-6 ,11,13	1-3,6,8-10, 12,13	1,2,5,6,8,9, 13	0,2,5,7,10,13	0,2,5,7,8,13	2,11,8,10,13	0,1,2,6,9,10, 13,14	0,1,9-11, 13,14[b]
Missing reported	For all	No	For 2 only	No	No	VACS Index	No	No	For 2,9,11	No	1
Missing>10%	10	-	2	-	-	VACS	-	-		-	1 (31%)
STATISTICS											
Missing data handling	Complete case	Complete case	Complete case	Complete case (8% [c])	Complete case	Multiple imputation	Complete case	Complete case	Complete case	Multiple imputation	Complete case
Model	Poisson	Cox	Logistic	Cox	Cox	Cox	Cox	Logistic	Cox	Poisson	Poisson
Variable selection	*A priori* & stepwise	*A priori* & stepwise	A priori, size/strength	*A priori*	Backwards stepwise	*A priori*	Backwards stepwise	Unclear	*A priori* & stepwise	*A priori*	*A priori*
Sensitivity	Yes	Yes	No report	Yes	No report	Yes	Yes	No report	No report	Yes	Yes

-Not reported VACS Index-A validated composite score which predicts all-cause mortality (age, viral load, haemoglobin, FIB-4, HCV, CD4, eGFR) [a] Abstracted from MR by trained staff (case definition not reported) but PPV value 90% (79-97) [b] Restricted to HCV-ves after ART start [c] 8% of observation deleted due to missing data Confounders: 0-age, 1-smoking, 2-BMI, 3-prior fracture, 4-parental hip fracture, 5-steroid use, 6-alcohol, 7-BMD/osteoporosis/bisphosphate use, 8-secondary osteoporosis, 9-ethnicity, 10-HCV/HBV 11-ART 12-menopause, 13-sex 14-other comorbidity

6.8.10 A comparison of findings across studies

There was a paucity of relevant publications, methods were highly variable and findings lacked agreement. Table 6.22a-d compares my findings to those of other studies before and after adjustment.

There was no evidence of an independent association between duration of HIV infection (only 2 studies) or duration of immune-suppression (only one study) and fracture incidence. There was also no evidence that higher current HIV viral load increased fracture risk [187, 195, 198]). Results across other exposures of interest were conflicting.

Table 6.22a: A comparison of my findings to those of other studies: The association between nadir CD4 cell count & fractures

				Nadir CD4 cell count			Factors accounted for in multivariable analysis*																
Study (Location)	# type	N/ Events	Point Estimate	Unit† Increase Cells/µL	Point Estimate (95% CI) p-value		TDF	PI	Mode	Smoking	Low BMD	BMI	HCV	Steroids	Opioids	Antacids	Alcohol	Hypertension	eGFR	Diabetes	Co-	HIV viral	Absolute CD4	Prior AIDS
					Univariable	Multivariable																		
My study (EU/ Canada)	All	13 243 266	IRR	100	**0.89** **(0.83-0.95)** **0.001**	0.99 (0.92-1.06) 0.8	**√**	**√**	√	x	x	x	**√**	x	x	x	x	x	x	x	x	**√**	**√**	**√**
SUN/ HOPS [183, 198] (US)	All	1 006 95	HR	100	0.98 (0.85-1.13)[c]	1.07 (0.90-1.27) 0.4	**√**	**√**	√	√	√	√	√	x	x	x	x	x	x	x	x	x	√	x
Boston [197] (US)	All	2 663 180	RR	<200 vs. ≥200	**3.7 (2.3-5.8)** **p<0.01**	**3.1 (1.9-5.0)** **<0.01**	**√**	x	x	x	√	x	√	√	x	x	√	x	x	x	√	x	x	x
WIHS [196] [186] (US)	All	1 713 300	HR	100	**0.91** **(0.84, 0.98)** **0.008**	Eliminated [a]	**√**	**√**	√	√	√	√	√	x	x	x	**√**	x	**√**	x	x	x	**√**	√
ALLRT [187] (US)	All	4 640 106	HR	50	0.98 (0.93-1.04) 0.6	Eliminated [a]	**√**	**√**	√	**√**	√	√	√	√	x	x	x	x	√	√	√	√	√	√
HOPS [182] (US)	All	5 826 233	HR	≥350 200-349 <200	1.00 1.31 (0.86–2.01) 0.2 **1.72 (1.20-2.48)** **0.004**	1.00 1.25 (0.8-1.92) 0.3 **1.60 (1.11-2.31)** **0.01**	√	√	x	**√**	x	√	√	x	x	√	**√**	x	x	√	√	√	√	√

*By restriction, matching, adjustment, randomisation or found not to be associated with the outcome and then dropped. All studies adjusted for or restricted/matched by age, sex, ethnicity (and cohort if applicable). Point estimates are in bold type if p<0.05. *Ticks in bold type if variables were time-updated †Units unless otherwise specified [a] Eliminated during backwards step-wise selection as p>0.05 [b] adjusted for ART exposure only (naïve/experienced/unknown) [c] No p reported [d] Co-morbidities: CVD, asthma, cancer, dementia, liver disease, rheumatoid arthritis, Celiac's disease, MS, depression, COPD, hyperparathyroidism and hyperthyroidism

Table 6.22b: A comparison of my findings to those of other studies: The association between current CD4 cell count & fractures

			Current CD4 cell count				Factors accounted for in multivariable analysis*																
Study (Location)	**Fracture type**	**N/n (HIV/ Event)**	**Point Estimate**	**Unit† Increase Cells/µL**	**Point Estimate (95% CI) p-value Univariable**	**Multivariable**	**TDF**	**PI**	**Mode**	**Smoking**	**Low BMD**	**BMI**	**HCV**	**Steroids**	**Opioids**	**Antacids**	**Alcohol**	**Hypertension**	**eGFR**	**Diabetes**	**Co-**	**HIV viral**	**Nadir CD4**	**Prior AIDS**
My study (EU/Canada)	All	13 243 266	IRR	100	1.02 (0.97-1.06) 0.5	1.02 (0.97-1.06) 0.5	**√**	**√**	√	x	x	x	√	x	x	x	x	x	x	x	x	**√**	**√**	**√**
SHCS [185] (Swiss)	Fragility	8 444 37	HR	√	-	**0.90 (0.85-0.95)[c]**	**x**	**x**	√	√	x	x	x	x	x	x	x	x	x	x	x	**√**	x	x
WIHS [196] [186] (US)	All	1 713 300	HR	100	0.97 (0.93,1.01) 0.1	Eliminated [a]	**√**	**√**	√	√	√	√	√	x	x	x	**√**	x	**√**	x	x	x	√	√
VACS-VS [132, 184] (US)	Fragility	40 115 588	HR	100	-	0.99 (0.95-1.02)[c]	**√**	**√**	√	√	x	√	√	√	x	√	√	x	**√**	x	**√**	**√**	x	x
Alfred Hospital [178] (Australia)	Fragility	183 [d] 73	OR	<200 vs.≥200[¥]	**6.77 (2.4-19.1) 0.001**	**4.91 (1.78-3.57) 0.002**	x	x	√	x	√	√	√	x	x	x	x	x	x	x	x	√	x	√

*By restriction, matching, adjustment, randomisation or found not to be associated with the outcome and then dropped**.** All studies adjusted for or restricted by age, sex, ethnicity (and cohort if applicable). - Not reported Point estimates are in bold type if p<0.05 * Ticks in bold type if variables were time-updated †Units unless otherwise specified ¥ Odds Ratio [a] Eliminated during backwards step-wise selection as p>0.05 [c] No p reported [d]61/122 (cases/controls) [e] Co-morbidities: CVD, asthma, cancer, dementia, liver disease, rheumatoid arthritis, Celiac's disease, MS, COPD, depression, hyperparathyroidism and hyperthyroidism

Table 6.22c: A comparison of my findings to those of other studies: The association between current HIV viral load & fractures

		Current HIV Viral Load					Factors accounted for in multivariable analysis*																	
Study	**Fracture type**	**N/n (HIV/ Events)**	**Point Estimate**	**Unit† Increase Cps/mL**	**Point Estimate (95% CI) p-value Univariable**	**Multivariable**	**TDF**	**PI**	**Mode**	**Smoking**	**Low BMD**	**BMI**	**HCV**	**Steroids**	**Opioids**	**Antacids**	**Alcohol**	**Hypertension**	**eGFR**	**Diabetes**	**Co-**	**Current CD4**	**Nadir CD4**	**Prior AIDS**
My study (EU/Canada)	All	13 243 266	IRR	Per log_{10}	**0.85 (0.77-0.93) <0.0001**	0.95 (0.85-1.05) 0.3	**√**	**√**	**√**	x	x	x	x	√	**x**	**x**	**x**	**x**	**x**	x	x	√	√	√
VACS-VS [132, 184] (US)	Fragility	40 115 588	HR	Per log_{10}	≠	**0.91 (0.88–0.94)≠**	**√**	**√**	√	√	x	√	√	√	x	√	√	x	**√**	x	**√**	**√**	x	x
Alfred Hospital [178] (Australia)	Fragility	183 73	OR	<400 vs ≥400	1.69 (0.97–1.32) 0.2	Eliminated [a]	x	x	√	x	√	√	√	x	x	x	x	x	x	x	x	√	x	√

*By restriction, matching, adjustment, randomisation or found not to be associated with the outcome and then dropped. All studies adjusted for age, sex, ethnicity (and cohort if applicable) ≠Not reported * ᶲNo p reported #Co-morbidities: CVD, asthma, cancer, dementia, liver disease, rheumatoid arthritis, Celiac's disease, MS, depression COPD, hyperparathyroidism and hyperthyroidism [a] Eliminated during backwards step-wise selection as p>0.05

Table 6.22d: A comparison of my findings to those of other studies: The association between prior AIDS and fractures

		Prior AIDS (Yes/No)				Factors accounted for in multivariable analysis*																
Study (Location)	**Fracture type**	**N/n (HIV/ Events)**	**Point Estimate**	**Point Estimate (95% CI) p-value Univariable**	**Multivariable**	**TDF**	**PI**	**Mode**	**Smoking**	**Low BMD**	**BMI**	**HCV**	**Steroids**	**Opioids**	**Antacids**	**Alcohol**	**Hypertension**	**eGFR**	**Diabetes**	**Co-morbidities**	**HIV viral**	**Absolute CD4**	**Nadir CD4**
My study (EU/Canada)	All	13 243 266	IRR	**1.90 (1.36-2.66) <0.0001**	**1.51 (1.07-2.12) 0.02**	**√**	**√**	√	x	x	x	√	x	x	x	x	x	x	x	x	**√**	**√**	√
Boston [197] (US)	All	2 663 180	RR	**3.7 (2.3-5.8)** <0.01	**3.1 (1.9-5.0) <0.01**	√	x	x	x	√	x	√	√	x	x	√	x	x	x	√	x	x	√
MACS [195] (US)	Fragility /All	1 221 182	IRR	-	1.3 (0.73-2.34) 0.4	**√**	**√**	x	**√**	√	√	√	x	x	x	√	√	√	√	x	√	√	x
DHCS [131] (Denmark)	Fragility /All	5 306 806	IRR	1.16 (0.84–1.58)≠	1.09 (0.77–1.55)≠	**x**	**x**	√	x	x	x	√	x	x	x	**x**	x	**x**	x	√	x	**√**	x
WIHS [196] [186] (US)	All (Fragility)	1 713 300	HR	**1.77 (1.41-2.22) <0.0001**	**1.57 (1.24-1.99) 0.0002**	**√**	**√**	√	√	√	√	√	x	x	x	**√**	x	**√**	x	x	x	**√**	√
HOPS [182] (US)	All	5 826 233	HR	**1.43 (1.07–1.92) 0.02**	Eliminated	√	√	x	**√**	x	√	√	x	x	√	**√**	x	x	√	√	√	√	√
ALLRT [187] (US)	All	4 640 106	HR	1.11 (0.70-1.78) 0.7	Eliminated	**√**	**√**	√	**√**	√	√	√	√	x	x	x	x	√	√	√	√	√	√

*By restriction, matching, adjustment, randomisation or found not to be associated with the outcome and then dropped**.** -Not reported All studies were adjusted for age, sex, ethnicity (and cohort if applicable) * Time-updated ticks in bold ≠No p reported ˚Co-morbidities: CVD, asthma, cancer, dementia, liver disease, rheumatoid arthritis, Celiac's disease, MS, COPD, depression, hyperparathyroidism and hyperthyroidism [a] Eliminated during backwards step-wise selection as p>0.05

7 Chapter 7: Myocardial Infarction Incidence

7.1 Introduction

Cardiovascular disease (CVD) is one of the leading causes of death in high-income countries [322]. Those with HIV are at greater risk of CVD including MI than those without the virus [212, 214, 323]. This could be due to HIV infection, ART or a different distribution of classic or novel CVD risk-factors in the HIV-positive population [212, 324-326].

In this chapter I present the results of the analysis I undertook to examine the associations between HIV-related factors and myocardial infarction (MI) incidence. In Section 7.2 I describe individuals included in this analysis and provide numbers and reasons for exclusion. I then go on to describe baseline characteristics (section 7.3) before presenting my main results (section 7.4 and 7.5). In section 7.6 I provide the results of my sensitivity analyses which is followed by a summary of my findings (Section 7.6). I then go on to compare my study characteristics and findings with those of other studies (Section 7.8).

7.2 Inclusion

In total 19 414 individuals with 121 first myocardial infarctions (MI), from the nine cohorts were potentially eligible for inclusion. Figure 7.1 shows numbers included and numbers and reasons for exclusion for both individuals and events.

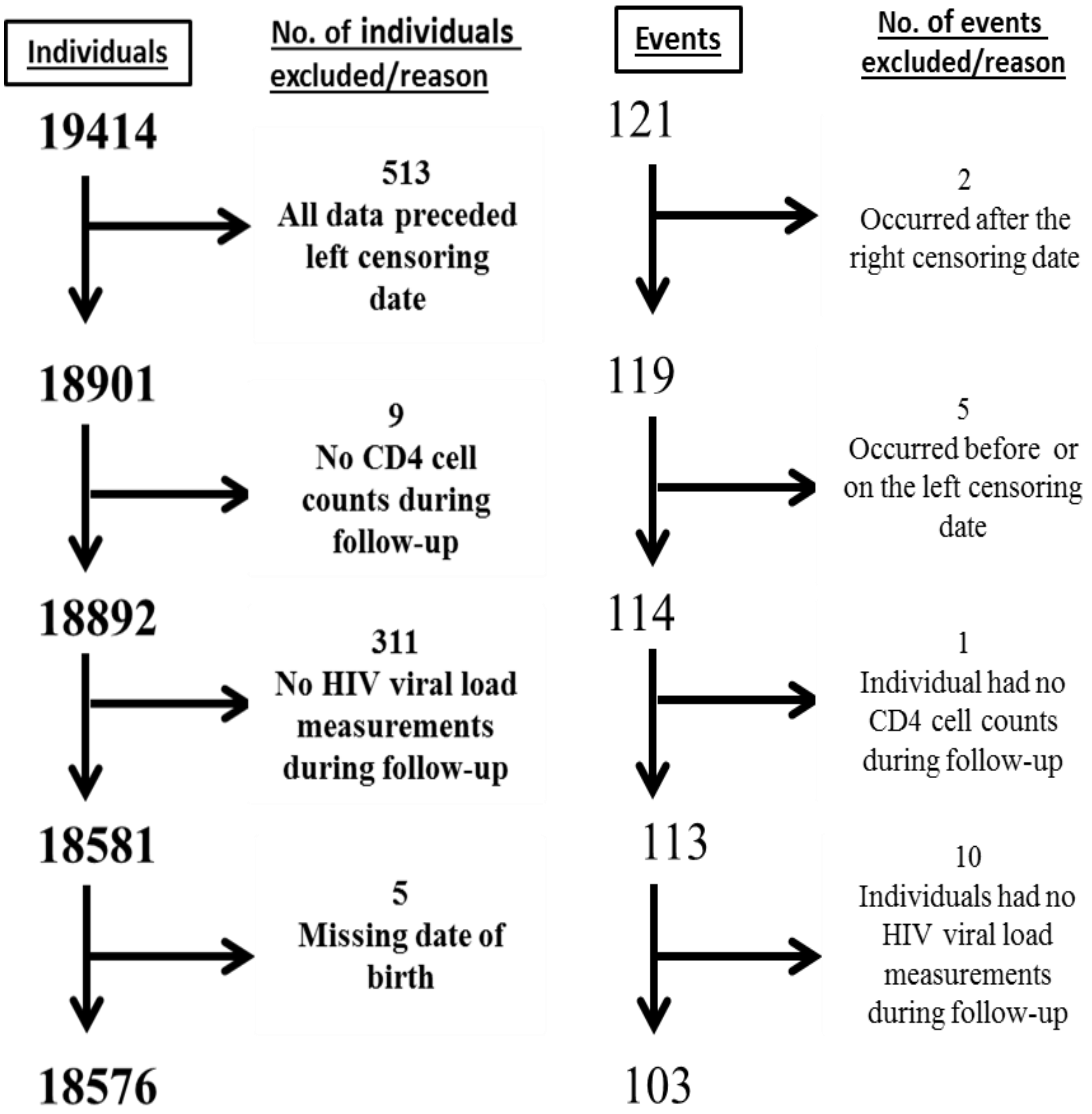


Figure 7.1: Numbers of HIV-positive individuals and events excluded from the MI analysis with reasons for their exclusion

In total 14.9% (18/121) of events and 4.3% (838/19 414) of individuals were excluded from the analysis (Figure 7.1). 18 576 people with a total follow-up of 128 654 person years (median 5.5 [2.4-10.7] years) were included; there were 103 first MI (event rate (per 1000 PYFU) of 0.8 (95%CI, 0.7-1.0)).

7.3 Baseline characteristics

Baseline characteristics of all individuals included in the MI analysis are shown in Table 7.1. Median age at seroconversion was 31.6 years with males being slightly older than females. The FHDH contributed the majority of individuals (57.3%). The risk for HIV acquisition in males was predominantly sex between men and in females was largely through heterosexual sex. A greater percentage of females were of black ethnicity (27.0%) than males (6.0%).

Table 7.1: Baseline characteristics of individuals in an analysis examining the association between duration/markers of HIV infection and MI incidence, stratified by sex

	Male (n=14 857)		Female (n=3 719)		Total (N=18 576)	
	Median [IQR] or n (%≠)					
Seroconversion age (years)	32.3 (26.9-39.2)		29.0 (24.1-35.7)		31.6 (26.2-38.6)	
Ethnicity						
Black	243	(4.0)	150	(14.9)	393	(5.5)
White	5 725	(93.0)	837	(83.0)	6 562	(91.6)
Other	186	(3.0)	22	(2.2)	208	(2.9)
Missing[1]	8 703		2 710		11 413	
Apparent ethnicity[2]						
Black	881	(6.0)	989	(27.0)	1870	(10.2)
White	13 484	(92.2)	2 622	(71.5)	16106	(88.0)
Other	264	(1.8)	54	(1.5)	326	(1.7)
Missing	228		54		282	
Mode						
MSM[#]/Bisexual	11 494	(79.8)	2	(0.1)	11 496	(64.1)
PWID~	690	(4.8)	406	(11.5)	1 096	(6.1)
Heterosexual	2 000	(13.9)	3 035	(86.1)	5 035	(28.1)
Other	224	(1.6)	81	(2.3)	305	(1.7)
Missing	449		195		644	
Cohort						
FHDH	8035	(54.1)	2615	(70.3)	10 650	(57.3)
AHIVCOS[3]	317	(2.1)	82	(2.2)	399	(2.2)
ICONA	1463	(9.9)	398	(10.7)	1861	(10.0)
NOR[4]	430	(2.9)	69	(1.9)	499	(2.7)
PRIMO	1142	(7.7)	200	(5.4)	1342	(7.2)
SEROCO	368	(2.5)	105	(2.8)	473	(2.6)
SAL[5]	169	(1.1)	23	(0.6)	192	(1.0)
UKR[6]	2538	(17.1)	195	(5.2)	2733	(14.7)
CoRIS	395	(2.6)	32	(0.9)	427	(2.3)

[1]Missing ethnicity also included prohibited. [2]Region of origin was used to determine likely ethnicity when ethnicity was missing. [3]AHIVCOS- Austrian HIV cohort [4]NOR-Oslo and Ulleval hospital cohort [5] SAL-Southern Alberta Clinic [6] UKR-UK Register of HIV Seroconverters [#]MSM, men who have sex with men ~PWID intravenous drug users ≠Percentage of non-missing total

7.4 Univariable analysis

7.4.1 Time-fixed variables

Table 7.2 shows the univariable analysis of time-fixed variables. MI rates were higher in individuals ≥40 vs <40 years at seroconversion. The highest rate was observed in those ≥60 years, which was nearly 5 times higher than in those <20 years. Confidence intervals were wide however, due to limited variability in age at seroconversion. Most follow-up (77.9%) was in those aged 20-<40 years at seroconversion.

Women were at lower risk of MI compared to men. Individuals of black ethnicity experienced lower rates of MI compared to those classified as white; there was an increased risk of MI in those of unknown ethnicity. No association was found between mode of transmission and MI rates, however rates of MI varied markedly by cohort. When compared to rates in the FHDH, which was the largest cohort, those from the Austrian HIV Cohort Study, the Olso and Ulleval Hospital Cohorts, SEROCO and the UK Register all experienced higher rates.

Table 7.2: The association between time-fixed factors and MI incidence in HIV-positive individuals: Rates and univariable analysis

Time-Fixed Factor	Event No	PYFU*	Rate (95%CI) (per 1000 PYFU)	IRR (95%CI)	p[1]
Seroconversion (age in years)					
<20	4	4 395	0.91 (0.34-2.42)	1	<0.0001
20-<40	60	100 180	0.60 (0.47-0.77)	0.66 (0.23-1.81)	
40-<60	32	22 531	1.42 (1.00-2.01)	1.56 (0.55-4.41)	
≥60	7	1 548	4.52 (2.15-9.49)	4.97 (1.45-17.0)	
Sex					
Male	91	99 860	0.91 (0.74-1.12)	1	
Female	12	28 794	0.42 (0.24-0.73)	0.46 (0.25-0.83)	0.01
Ethnicity					
White	56	47 113	1.19 (0.91-1.54)	1	0.001
Black	1	2 474	0.40 (0.05-2.87)	0.34 (0.05-2.46)	
Other	0	1 090	-	-	
Missing	46	77 978	0.59 (0.44-0.78)	0.50 (0.34-0.73)	
Apparent ethnicity[2]					
White	94	113 312	0.83 (0.68-1.02)	1	0.4
Other	1	1 726	0.58 (0.08-4.11)	0.70 (0.10-5.01)	
Black	1	10 821	0.09 (0.01-0.66)	0.11 (0.02-0.80)	
Missing	7	2 796	2.50 (1.19-5.25)	3.02 (1.40-6.50)	
Mode					
MSM[#]	62	76 016	0.82 (0.64-1.05)	1	0.9
Heterosexual	26	36 548	0.71 (0.48-1.04)	0.87 (0.55-1.38)	
PWID~	8	9 902	0.81 (0.40-1.62)	0.99 (0.47-2.07)	
Other	2	2 444	0.82 (0.20-3.27)	1.00 (0.25-4.10)	
Missing	5	3 743	3.74 0.56-3.21)	1.64 (0.66-4.07)	
Cohort					
FHDH	38	73 961	0.51 (0.37-0.71)	1	<0.0001
AHIVCOS[3]	6	3 062	1.96 (0.88-4.36)	4.30 (1.92-9.62)	
ICONA	7	9 984	0.70 (0.33-1.47)	1.33 (0.59-2.97)	
NOR[4]	18	4 511	3.99 (2.51-6.33)	10.7 (6.57-17.5)	
PRIMO	1	7 431	0.13 (0.02-0.95)	0.25 (0.03-1.85)	
SEROCO	8	5 269	1.52 (0.76-3.04)	2.85 (1.33-6.09)	
SAL[5]	1	865	1.16 (0.16-8.21)	2.19 (0.30-16.0)	
UKR[6]	24	22 623	1.06 (0.71-1.58)	2.17 (1.32-3.57)	
CoRIS	0	948	-	-	

[#]MSM-Men who have sex with men ~PWID-Intravenous drug users [1]Test for heterogeneity excluding missing [2] Region of origin was used as a guide for ethnicity when ethnicity was missing. [3]AHIVCOS- Austrian HIV cohort [4]NOR-Oslo and Ulleval hospital cohort [5]SAL-Southern Alberta hospital cohort [6]UKR-UK Register of HIV seroconverters*PYFU-person-years of follow-up

7.4.2 Time-updated variables

Table 7.3 shows results of the univariable analysis of time-updated factors. Duration of infection was associated with MI incidence, with increasing MI rates with longer duration of infection. Rates were more than nine times higher in those who had been infected >12 years when compared to those infected ≤3 years.

Current age was also strongly associated with MI, with rates for those >60 years of age being about thirteen times higher than those in the ≤40 years group. There was a borderline association (p=0.06) between current CD4 cell count and MI incidence; individuals with higher current CD4 cell counts generally experienced lower rates. Increasing nadir CD4 cell count was associated with decreasing rates of MI. No association was apparent between $\log_{10}$ HIV viral load and MI.

There was no evidence that current abacavir or duration of indinavir exposure were associated with MI, but MI incidence increased with increasing duration of exposure to lopinavir. MI rates increased with longer duration of immunosuppression ≤200/≤100 cells/µL with those having >12 months of exposure at either level experiencing more than a 2.5-fold higher rate. Few patients experienced sustained immunosuppression ≤50cells/µL, but those who did had elevated rates of MI. An association between HCV infection and MI was of borderline significance (p=0.10). There was evidence that those who had had a prior AIDS event were at increased risk of MI compared to those without one. A calendar period effect was noted, with much higher rates of MI after 2000.

Table 7.3: The association between time-updated factors and MI incidence in HIV-positive individuals: Rates and univariable analysis (continues)

	Event (n)	PYFU*	Rate/1000 PYFU (95%CI)	IRR (95%CI)	p[1]
Duration of Infection					
≤3 years	7	31890	0.22 (0.10-0.46)	1	<0.0001
>3-6	16	30981	0.52 (0.32-0.84)	2.35 (0.97-5.72)	
>6-9	18	23526	0.77 (0.48-1.21)	3.49 (1.46-8.35)	
>9-12	10	17051	0.59 (0.32-1.09)	2.67 (1.02-7.02)	
>12	52	25207	2.06 (1.57-2.71)	9.40 (4.27-20.7)	
Current age					
≤40 years	19	71018	0.27 (0.17-0.42)	1	<0.0001
>40-50	36	38859	0.93 (0.67-1.28)	3.46 (1.99-6.04)	
>50-60	31	13986	2.22 (1.56-3.15)	8.29 (4.68-14.7)	
>60	17	4792	3.55 (2.20-5.71)	13.26 (6.9-25.5)	
Current CD4 (cells/µl)					
≤100	5	2569	1.95 (0.81-4.68)	2.34 (0.94-5.84)	0.06
101-200	5	5207	0.96 (0.40-2.30)	1.15 (0.46-2.88)	
201-350	21	19781	1.06 (0.69-1.63)	1.28 (0.77-2.11)	
351-500	16	32218	0.50 (0.30-0.81)	0.60 (0.34-1.04)	
>500	56	67279	0.83 (0.64-1.08)	1	
Missing	0	1600	-	-	
Nadir CD4 cell count (cells/µl)					
≤100	24	14671	1.64 (1.10-2.44)	1	0.0004
101-200	23	21192	1.09 (0.72-1.63)	0.66 (0.37-1.18)	
201-350	34	43811	0.78 (0.55-1.09)	0.47 (0.28-0.80)	
351-500	8	25943	0.31 (0.15-0.62)	0.19 (0.08-0.42)	
>500	14	21438	0.65 (0.39-1.10)	0.40 (0.21-0.77)	
Missing	0	1600	-	-	
Current viral load (copies/mL)					
<1000	70	72943	0.96 (0.76-1.21)	1	0.40
1000-9999	13	16515	0.77 (0.45-1.34)	0.82 (0.45-1.47)	
10000-99999	12	21667	0.55 (0.31-0.96)	0.57 (0.31-1.05)	
≥100000	8	7295	1.08 (0.54-2.17)	1.13 (0.55-2.36)	
Missing	0	9368	-	-	
Current ABC[2]					
Off	86	107654	0.80 (0.65-0.99)	1	1.00
On	17	21000	0.81 (0.50-1.30)	1.01(0.60-1.70)	
Years of IDV[3] exposure					
Never exposed	87	118562	0.73 (0.59-0.91)	1	0.7
≤1	8	5293	1.51 (0.76-3.02)	2.06 (1.00-4.24)	
>1-2	4	2022	1.98 (0.74-5.27)	2.70 (0.99-7.35)	
>2-3	2	1165	1.72 (0.43-6.86)	2.34 (0.57-9.50)	
>3	2	1612	1.24 (0.31-4.96)	1.69 (0.42-6.87)	
Per additional Year	-	-	-	1.14 (0.97-1.35)	0.1

[1] Test for heterogeneity with missing excluded [2]ABC-Abacavir [3]Indinavir *PYFU-person-years of follow-up

Table 7.3 (continued): The association between time-varying factors and MI incidence in HIV-positive individuals: Rates and univariable analysis

	Event (n)	PYFU*	Rate/1000 PYFU (95%CI)	IRR (95%CI)	p[1]
Years of LOP[2] exposure					
Never exposed	86	118 301	0.73 (0.59-0.89)	1	0.006
≤2	6	4 542	1.32 (0.59-2.94)	1.82 (0.79-4.16)	
2-4	2	2 076	0.96 (0.24-3.85)	1.33 (0.33-5.39)	
4-6	2	1 314	1.52 (0.38-6.09)	2.09 (0.52-8.51)	
>6	7	2 422	2.90 (1.38-6.06)	3.98 (1.84-8.59)	
Per additional year	-	-	-	1.15 (1.07-1.23)	<0.0001
Duration of immune suppression ≤200cells/µL					
CD4 always >200	56	91 367	0.61 (0.47-0.79)	1	0.007
>0-6 months	15	15 849	0.95 (0.57-1.57)	1.54 (0.87-2.73)	
>6-12 months	9	5 782	1.56 (0.81-2.99)	2.54 (1.26-5.13)	
>12 months	23	14 055	1.63 (1.09-2.46)	2.67 (1.64-4.34)	
Missing	0	1 600	-	-	
Duration of immune suppression ≤100cells/µL					
CD4 always >100	79	112 455	0.70 (0.56-0.88)	1	0.002
>0-6 months	9	7 131	1.26 (0.66-2.43)	1.80 (0.90-3.58)	
>6-12 months	5	2 458	2.03 (0.85-4.89)	2.90 (1.17-7.15)	
>12 months	10	5 011	2.00 (1.07-3.71)	2.84 (1.47-5.48)	
Missing	0	1 600	-	-	
Duration of immune suppression ≤50cells/µL					
CD4 always >50	86	118 975	0.72 (0.58-0.89)	1	<0.0001
>0-6 months	12	4 177	2.87 (1.63-5.06)	3.97 (2.17-7.27)	
>6 months	5	3 901	1.28 (0.53-3.08)	1.77 (0.72-4.37)	
Missing	0	1 600	-	-	
Prior HCV					
No	88	115 556	0.76 (0.62-0.94)	1	0.1
Yes	15	13 098	1.15 (0.69-1.90)	1.50 (0.87-2.60)	
Prior AIDS					
No	79	116 537	0.68 (0.54-0.85)	1	<0.0001
Yes	24	12 117	1.98 (1.33-2.96)	2.92 (1.85-4.61)	
Calendar Period					
≤2000	4	20 776	0.19 (0.07-0.51)	0.18 (0.06-0.49)	0.007
>2000-2005	28	33 810	0.83 (0.57-1.20)	0.76 (0.48-1.21)	
>2005-2010	51	47 089	1.08 (0.82-1.43)	1	
>2010	20	26 979	0.74 (0.48-1.15)	2.11 (1.32-11.3)	
Calendar Period					
≤2000	4	20 776	0.19 (0.07-0.51)	1	0.002
>2000	99	107 878	0.92 (0.75-1.12)	4.77 (1.75-12.95)	

[1]Test for heterogeneity with missing excluded [2]LOP-Lopinavir *PYFU-person-years of follow-up

Table 7.4 shows IRRs for the univariable analysis of continuous covariates. For each additional 10 years of HIV infection duration MI rates more than trebled. There was also evidence that MI rates approximately doubled for each additional decade of current age, and for each additional decade of age at seroconversion rates were increased about 60%. Nadir CD4 was associated with MI incidence; those with higher nadirs experienced reduced rates. There was no evidence of an association between either current CD4 cell count or current HIV viral load and MI incidence. Duration of immune-suppression at both ≤200 and ≤100 cells/μL were positively associated with MI incidence, but duration ≤50 cells/μL was not (although few individuals experienced such profound immune-suppression).

Both age at seroconversion and current age were strongly positively correlated (Pearson's correlation coefficient (r_P) =0.91). Age at seroconversion was weakly negatively correlated with duration of HIV infection (r_P =-0.10). Current age was moderately positively correlated with duration of HIV infection (r_P =0.36). Current age was included in preference to age at seroconversion in subsequent models as it was more strongly associated with MI incidence, is a well-known predictor of MI and the two variables could not both be included (Section 4.4.2.3).

Although there was evidence of increasing MI rates with later calendar year, the association was found not to be linear and was best described by a binary variable (based on AIC scores) which were used in the multivariable analysis.

In multivariable analysis, *a priori* potential confounders identified from the literature review (see section 2.4.4) and available in the CASACADE 2013 dataset comprised: current age, sex, duration of lopinavir, duration of indinavir, current abacavir and calendar period. Adjustment for cohort and calendar period was also predetermined.

Table 7.4: The association between continuous covariates and MI incidence in HIV-positive individuals: Univariable analysis*

Time-updated covariate	IRR (95%CI)	p
Duration of HIV infection~	3.51 (2.59-4.77)	<0.0001
Seroconversion age~	1.63 (1.38-1.92)	<0.0001
Current age~	2.12 (1.83-2.47)	<0.0001
Nadir CD4 cell count (per 100 cell/μL increase)	0.82 (0.73-0.92)	0.001
Current HIV Viral Load (per log_{10} copy increase)	0.89 (0.77-1.03)	0.1
Current CD4 cell count (per 100 cell/μL increase)	1.00 (0.99-1.02)	0.4
Duration of Immunosuppression¥ (≤200 cells/μL)	1.18 (1.10-1.27)	<0.0001
Duration of Immunosuppression¥ (≤100 cells/μL)	1.22 (1.06-1.42)	0.006
Duration of Immunosuppression¥ (≤50 cells/μL)	1.16 (0.85-1.59)	0.4
Duration of Lopinavir¥	1.15 (1.07-1.23)	<0.0001
Duration of Indinavir¥	1.14 (0.97-1.35)	0.1
Calendar year¥	1.06 (1.01-1.11)	0.009

*All variables were time-updated except age at seroconversion ~per additional 10 years
¥per additional year

7.5 Multivariable analysis

In the basic multivariable model (A0, Table 7.5) there was evidence for an association between current age, cohort and calendar period and MI incidence. There was no evidence of a statistically significant association ($p<0.05$) between the three ART variables or sex and MI. Duration of lopinavir had been found to be associated with MI in the univariable analysis, but this association was attenuated after adjustment for current age. All variables in model A0 were then tested for interaction with current age and sex. Unlike the analyses of fractures or diabetes, ethnicity was not included as there was little evidence to support an association, either in the literature or in my univariable analysis. Table 7.6 shows the interaction term p values, none of which were found to statistically significant ($p<0.1$).

Table 7.5: The association between demographic factors/potential confounders and MI incidence in HIV-positive individuals: Basic multivariable model A0

Variable	IRR (95%CI)	p
Current age~	2.00 (1.70-2.35)	<0.0001
Sex		
Male	1	0.2
Female	0.64 (0.35-1.18)	
Cohort		
FRENCH[1]	1	<0.0001
AHIVCOS[2]	3.19 (1.35-7.53)	
NOR[3]	6.56 (3.78-11.4)	
SEROCO/SAL[4]	3.92 (1.86-8.23)	
UKR[5]	1.76 (1.05-2.95)	
Duration of Lopinavir¥	1.05 (0.97-1.14)	0.2
Duration of Indinavir¥	1.06 (0.88-1.27)	0.5
Current Abacavir		
Off	1	0.5
On	0.85 (0.50-1.44)	
Calendar Period		
≤2000	1	0.01
>2000	3.78 (1.35-10.5)	

[1]French-comprised FHDH/ICONA/PRIMO/CoRIS [2]AHIVCOS-Austrian HIV cohort [3]NOR-Oslo and Ulleval Hospital Cohort [4]Southern Alberta Clinic ~per additional 10 years ¥per additional year

There was no evidence of interaction (Table 7.6). Therefore model A was identical to model A0 (shown in Table 7.5).

Table 7.6: The assessment of interaction between sex, age and ethnicity and other variables in the basic multivariable model A0
(which examines the association between demographic factors/potential confounders and MI incidence in HIV-positive individuals)

Variable	Interaction term p values*	
	Current age	Sex
Sex	0.4	x
Cohort	0.3	0.4
Indinavir¥	0.7	0.7
Lopinavir¥	0.7	0.3
Abacavir (Yes/No)	0.2	0.4
Calendar period (≤2000/>2000)	0.4	a

*Interaction terms were added individually to model A0 and tested for significance. a-could not be calculated as no events occurred in some categories ¥ per additional year

Table 7.7 shows the results for each covariate of interest added separately to model A. Duration of HIV infection remained statistically significantly associated with the outcome after adjustment. Duration of immune suppression ≤200 cells/µL and prior AIDS also remained statistically significant (p<0.1) and were further considered for inclusion in the final model. Duration of immune suppression ≤100 cells/µL was of borderline significance (p=0.1) and collinear with duration of immune suppression ≤200 cells/µL so it was not considered further. When duration of HIV infection, duration of immune suppression ≤200 cells and prior AIDS were all added to model A, duration of immune suppression was no longer significant (p=0.4) and was removed. Duration of HIV infection and prior AIDS remained significant at p<0.1 and were therefore retained.

The final model (model J) is shown in Table 7.8. There was no evidence that duration of HIV infection of prior AIDS interacted with either current age or sex in the final model.

Table 7.7: The association between duration/markers of HIV infection and MI incidence in HIV-positive individuals: IRRs for each covariate of interest added separately to multivariable model A (Model B_1-I_1) with univariable results shown for comparison

Model*	Covariate	Multivariable IRR (95%CI)*	p	Univariable IRR (95%CI)	p
B_1	**Duration of HIV Infection~**	2.02 (1.44-2.83)	<0.0001	3.51 (2.59-4.77)	<0.0001
C_1	**Current CD4 cell count (per 100 cell/μL increase)**	1.00 (0.99-1.02)	0.3	1.00 (0.99-1.02)	0.4
D_1	**Log viral load (per log_{10} copy/mL increase)**	1.06 (0.92-1.22)	0.4	0.89 (0.77-1.03)	0.1
E_1	**Nadir CD4 cell count (per 100 cell increase)**	0.92 (0.47-1.37)	0.2	0.82 (0.73-0.92)	0.001
F_1	**Years of immune suppression¥ (≤200 cells/μL)**	1.09 (1.01-1.19)	0.04	1.18 (1.10-1.27)	<0.0001
G_1	**Years of immune suppression¥ (≤100 cells/μL)**	1.12 (0.97-1.32)	0.1	1.22 (1.06-1.42)	0.006
H_1	**Years of immune suppression¥ (≤50 cells/μL)**	1.05 (0.75-1.47)	0.8	1.16 (0.85-1.59)	0.4
I_1	**Prior AIDS (yes vs. no)**	2.10 (1.31-3.36)	0.002	2.92 (1.85-4.61)	<0.0001

*Model A comprised: current age; sex; cohort; years of lopinavir and indinavir exposure; current abacavir exposure and binary calendar period ¥per additional year ~per additional 10 years

Table 7.8 shows the final multivariable model for the main MI analysis (model J). Each additional 10 years of HIV infection was associated with a near doubling of MI incidence after adjustment for current age and other factors. Having had AIDS increased MI rates by over 70%.

Table 7.8: The association between duration/markers of HIV infection and MI incidence in HIV-positive individuals after adjusting for demographic factors/potential confounders: The final multivariable model (model J)

Variable	IRR (95%CI)	p
Duration of infection~	1.87 (1.32-2.64)	<0.0001
Current age~	1.85 (1.55-2.20)	<0.0001
Sex		
Male	1	
Female	0.60 (0.33-1.11)	0.10
Cohort		
FRENCH[1]	1	
AHIVCOS[2]	3.56 (1.50-8.44)	<0.0001
NOR[3]	5.86 (3.35-10.2)	
SEROCO/SAL[4]	3.15 (1.50-6.64)	
UKR[5]	1.54 (0.92-2.60)	
Years of Lopinavir¥	1.00 (0.92-1.08)	1.00
Years of Indinavir¥	1.00 (0.82-1.21)	1.00
Current Abacavir		
No	1	
Yes	0.73 (0.43-1.24)	0.25
Prior AIDS		
No	1	
Yes	1.72 (1.06-2.78)	0.03
Calendar period		
≤2000	1	0.04
>2000	2.90 (1.03-8.17)	

[1]FRENCH-FHDH/ICONA/PRIMO/CoRIS [2] AHIVCOS-Austrian HIV cohort [3] NOR-Oslo and Ulleval hospital cohort [4]SAL-Southern Alberta Clinic [5] UKR-UK Register of HIV Seroconverters ¥per additional year ~per additional 10 years

7.6 Sensitivity analyses

I undertook a series of sensitivity analysis s to test the robustness of assumptions.

Results of sensitivity analysis 1a are shown in Table 7.9. This sensitivity analysis investigated the effects of censoring follow-up after the first gap in CD4 measurements of at least one year. These results were similar to those found in the main analysis, but duration of immune suppression ≤200 cells/μL was no longer statistically significantly associated with MI, likely due to a reduction in power. No covariates derived from CD4 cell counts were eligible for inclusion in the final model.

Table 7.9: Sensitivity analysis 1a: Each CD4 related covariate added separately (models C/E-G) to model A after censoring follow-up after the first gap in CD4 measurements of ≥1 year (with main analysis results for comparison)

Sensitivity model	Covariate	Main analysis IRR (95%CI)*	p	Sensitivity model 1_a IRR (95%CI)*	p
C_{1a}	**Current CD4 cell count (per 100 cell/μL increase)**	1.00 (0.99-1.02)	0.3	1.00 (0.99-1.02)	0.2
E_{1a}	**Nadir CD4 cell count (per 100 cell/μL increase)**	0.92 (0.47-1.37)	0.2	1.00 (0.92-1.10)	0.8
F_{1a}	**Duration of immune suppression¥ (≤200cells/μL)**	1.09 (1.01-1.19)	0.04	1.07 (0.94-1.22)	0.3
G_{1a}	**Duration of immune suppression¥ (≤100cells/μL)**	1.12 (0.97-1.32)	0.1	1.15 (0.89-1.49)	0.3
H_{1a}	**Duration of immune suppression¥ (≤50cells/μL)**	1.05 (0.75-1.47)	0.8	1.17 (0.80-1.72)	0.4

*Model A comprised: current age, sex, cohort, years of lopinavir and indinavir exposure, current abacavir exposure and binary calendar period ¥per additional year

Table 7.10 shows the results for sensitivity analysis 1b. This investigated the effects of censoring follow-up after the first gap in CD4 measurements of at least one year but allowing individuals to re-enter the analysis if subsequent CD4 cell counts became available. Again results of multivariable analysis were similar to those found in the main analysis, but duration of immune suppression ≤200 cells/μL was of borderline statistical significance (p=0.09). When it was added to Model A in addition to either prior AIDS (IRR per additional year of immune-suppression, 1.01; 0.82-1.24; p=1.0), duration of HIV infection (IRR per additional year, 1.03; 0.93-1.14; p=0.6) or both (0.95; 0.76-1.20; p=0.7) there was no evidence of an association with MI incidence and duration of immune suppression ≤200 cells/μL was not included in the final model.

Table 7.10: Sensitivity analysis 1b: Each CD4 related covariate added separately (models C/E-G) to model A after censoring follow-up after the first gap in CD4 measurements of ≥1 year, but allowing re-entry to follow-up if subsequent CD4 measurements were recorded (with main analysis results for comparison)

Sensitivity model	Covariate	Main analysis IRR (95%CI)*	p	Sensitivity model 1_b IRR (95%CI)*	p
C_{1b}	Current CD4 cell count (per 100 cell/μL increase)	1.00 (0.99-1.02)	0.3	1.00 (0.99-1.02)	0.2
E_{1b}	Nadir CD4 cell count (per 100 cell/μL increase)	0.92 (0.47-1.37)	0.2	0.92 (0.82-1.03)	0.2
F_{1b}	Duration of immune suppression¥ (≤200cells/μL)	1.09 (1.01-1.19)	0.04	1.09 (0.99-1.19)	0.09
G_{1b}	Duration of immune suppression¥ (≤100cells/μL)	1.12 (0.97-1.32)	0.1	1.10 (0.91-1.33)	0.3
H_{1b}	Duration of immune suppression¥ (≤50cells/μL)	1.05 (0.75-1.47)	0.8	1.08 (0.76-1.54)	0.6

*Model A comprised: current age, sex, cohort, years of lopinavir and indinavir exposure, current abacavir exposure and binary calendar period ¥per additional year

Sensitivity analyses assessing the impact of carrying HIV viral loads forward for more than a year are shown in Table 7.11 (sensitivity analysis 2a) and Table 7.12 (sensitivity analysis 2b); these results were more similar to the univariable results from the main analysis, than main analysis results after adjustment. Log_{10} HIV viral load remained, however, ineligible for consideration for inclusion in the final model due to lack of evidence for a statistically significant association with MI ($p<0.1$).

Table 7.11: Sensitivity analysis 2a: HIV viral load added to model A after censoring follow-up after the first gap in HIV viral load measurements of ≥1 year (with main analysis results for comparison)

Sensitivity model	Covariate	Main analysis IRR (95%CI)*	p	Sensitivity model 2_a IRR (95%CI)*	p
D_{2a}	**Log_{10} HIV viral load (per log_{10} copies/mL increase)**	1.06 (0.92-1.22)	0.4	0.87 (0.73-1.04)	>0.1

*Model A comprised: current age, sex, cohort, years of lopinavir and indinavir exposure, current abacavir exposure and binary calendar period

Table 7.12: Sensitivity analysis 2b: HIV viral load added to model A after censoring follow-up after the first gap in HIV viral load measurements of ≥1 year, but allowing re-entry if subsequent HIV viral load measurements were recorded (with main analysis results)

Sensitivity model	Covariate	Main analysis IRR (95%CI)*	p	Sensitivity model 2_b IRR (95%CI)*	p
D_{2b}	**Log_{10} HIV viral load (per log_{10} copies/mL increase)**	1.06 (0.92-1.22)	0.4	0.91 (0.79-1.06)	0.2

*Model A comprised: current age, sex, cohort, years of lopinavir and indinavir exposure, current abacavir exposure and binary calendar period

Sensitivity analysis 3 restricted the analysis to follow-up ≥01/01/2005. There were 70 MIs and 16 438 individuals included in the analysis, so 33 events and 2 138 individuals from the main analysis were excluded. Total PYFU was approximately 57% of that of the main analysis (73 483 years compared to 128 654) and the crude first MI rate was 1.0 (0.8-1.2) per 1000 PYFU (vs. 0.8 (0.7-1.0) in the main analysis). The directions of all associations in the univariable analysis were the same as in the main analysis and the effect size was similar.

Table 7.13 shows the results of adding each covariate of interest to model A (minus adjustment for calendar period; models B_3-I_3) and compares these results to those found in the main analysis. Unlike the main analysis, nadir CD4 cell count and duration of immune suppression at ≤100 cells/µL were statistically significantly associated with MI incidence ($p<0.1$).

Table 7.13: Sensitivity analysis 3: Each covariate of interest added separately to multivariable model A (models B_3-I_3) after restricting follow-up to ≥01/01/2005 (with main analysis results for comparison)

Sensitivity model	Covariate	Main analysis IRR (95%CI)*	p	Sensitivity model 3 IRR (95%CI)*	p
B_3	Duration of HIV Infection~	2.02 (1.44-2.83)	<0.0001	2.24 (1.52-3.31)	<0.0001
C_3	Current CD4 cell count (per 100 cell/µL increase)	1.00 (0.99-1.02)	0.3	1.00 (0.99-1.02)	0.5
D_3	Log viral load (per $\log_{10}$ copies/mL increase)	1.06 (0.92-1.22)	0.4	1.06 (0.89-1.27)	0.5
E_3	Nadir CD4 cell count (per 100 cell/µL increase)	0.92 (0.47-1.37)	0.2	0.88 (0.75-1.02)	0.09
F_3	Duration of immune suppression¥ (≤200cells/µL)	1.09 (1.01-1.19)	0.04	1.14 (1.05-1.25)	0.003
G_3	Duration of immune suppression¥ (≤100cells/µL)	1.12 (0.97-1.32)	0.1	1.20 (1.02-1.41)	0.03
H_3	Duration of immune suppression¥ (≤50cells/µL)	1.05 (0.75-1.47)	0.8	1.08 (0.71-1.64)	0.7
I_3	Prior AIDS (yes vs no)	2.10 (1.31-3.36)	0.002	2.35 (1.34-4.13)	0.003

*Model A comprised: current age, sex, cohort, years of lopinavir and indinavir exposure, current abacavir exposure ~per additional 10 years ¥per additional year

Nadir CD4 cell count, duration of immune suppression ≤200 cells/μL and prior AIDS were then included in a model with all variables in model A and duration of HIV infection. A backwards selection process was then undertaken, in which nadir CD4 cell count, duration of immune suppression ≤200 cells/μL and prior AIDS were the only variables permitted to be removed from the model ($p<0.1$). Nadir CD4 cell count was removed first ($p=0.9$), after refitting the model duration of immune suppression ≤200 cells/μL was removed ($p=0.3$). After the model was refit prior AIDS was retained ($p=0.04$) and the process stopped. The final model, in-line with the main analysis, only retained two covariates of interest, duration of HIV infection 2.06 (1.38-3.08; $p<0.0001$) and prior AIDS 1.81 (1.02-3.22; $p=0.04$).

In sensitivity analysis 4 (Table 7.14), due to some uncertainty about when systematic data capture on MIs started for each cohort, the first possible date of entry to follow-up for each cohort was changed to the day after the first recorded MI event for that cohort. There were 98 MIs and 17 794 individuals included in the analysis, so 5 events and 782 individuals from the main analysis were excluded. Total PYFU was approximately 83% of that of the main analysis (107 019 years compared to 128 654) and the crude first MI rate was 0.9 (0.8-1.1) per 1000 PYFU (vs. 0.8 (0.7-1.0) in the main analysis).

Table 7.14 shows each covariate of interest added to multivariable model A. All associations were very similar with respect to size, direction and statistical significance when compared to the main analysis. Prior AIDS and duration of immune suppression ≤200 cells/μL were included in a model with all model A variables and duration of HIV infection. Backwards selection was then undertaken in which only prior AIDS and duration of immune suppression were permitted to be dropped. Duration of immune suppression was removed ($p=0.8$). When the model was re-run both duration of HIV infection (1.88; 1.32-2.68; $p<0.0001$) and prior AIDS (1.78; 1.09-2.88; $p=0.02$) were statistically significantly associated with MI incidence after adjustment for current age and other factors.

Table 7.14: Sensitivity analysis 4: Each covariate of interest added separately to multivariable model A (models B_4-I_4) after restricting entry to follow-up by cohort to the day after the first MI was recorded for that cohort (main analysis results included for comparison)

Sensitivity model	Covariate	Main analysis IRR (95%CI)*	p	Sensitivity model 4 IRR (95%CI)*	p
B_4	Duration of HIV Infection~	2.02 (1.44-2.83)	<0.0001	2.04 (1.45-2.88)	<0.0001
C_4	Current CD4 cell count (per 100 cell/µL increase)	1.00 (0.99-1.02)	0.3	1.01 (0.99-1.02)	0.4
D_4	Log viral load (per log_{10} copies/mL increase)	1.06 (0.92-1.22)	0.4	1.07 (0.93-1.24)	0.4
E_4	Nadir CD4 cell count (per 100 cell/µL increase)	0.92 (0.47-1.37)	0.2	0.91 (0.81-1.03)	0.2
F_4	Duration of immune suppression¥ (≤200cells/µL)	1.09 (1.01-1.19)	0.04	1.10 (1.01-1.19)	0.03
G_4	Duration of immune suppression¥ (≤100cells/µL)	1.12 (0.97-1.32)	0.1	1.13 (0.97-1.32)	>0.1
H_4	Duration of immune suppression¥ (≤50cells/µL)	1.05 (0.75-1.47)	0.8	1.05 (0.75-1.48)	0.8
I_4	Prior AIDS (yes vs no)	2.10 (1.31-3.36)	0.002	2.16 (1.35-3.48)	0.001

*Model A comprised: current age; sex; cohort; years of lopinavir and indinavir exposure; current abacavir exposure and binary calendar period ~per additional 10 years ¥per additional year

The next two sensitivity analyses used multivariable fractional polynomials (mfp) to allow for non-linear associations in the model. Sensitivity analysis 5 explored possible non-linear associations in model J, the final model from the main analysis.

Table 7.15 compares the inclusion of the best fitting FP2 transformation (see Table 4.5-1 for all transformations considered) to the variable in its linear form for the four continuous variables included in the final multivariable MI model (model J). For duration of HIV infection the best-fitting FP2 transformation was found to be a $1/x^2$ transformation for p^1 and linear (no transformation) for p^2 (where FP2= $\beta_1 X^{p1}+\beta_2 X^{p2}$, see Section 4.5.5); there was no evidence that the FP2 transformation was a better fit than the linear term (p=0.1). For current age, the best fitting FP2 transformation was $1/x^2$ for both p^1 and p^2, but again there was no evidence it improved the model (p=0.5). The linear form of the variable was also chosen for both years on indinavir and years on lopinavir (where p^1 and p^2 of the best fitting FP2 were $1/x^2$ and $1/\sqrt{x}$).

Table 7.15: Sensitivity analysis 5: Assessing fractional transformations (mfp) of variables included in the final multivariable analysis (model J) on the association between duration/markers of HIV infection and MI*

Variable	Model	vs.	Deviance	Deviance difference	p	Powers	vs.
Duration of Infection~	Linear	FP2	2220	5.6	0.1	1	-2 1
	Final		2220			1	
Current age~	Linear	FP2	2220	2.6	0.5	1	-2 -2
	Final		2220			1	
Years on Indinavir¥	Linear	FP2	2220	1.6	0.7	1	-2 -2
	Final		2220			1	
Years on Lopinavir¥	Linear	FP2	2220	6.6	0.2	1	-2 -0.5
	Final		2220			1	

~Per additional 10 years ¥ per additional year *Also adjusted for: sex; cohort; current abacavir; prior AIDS and calendar period

Sensitivity analysis 6 used mfp and backwards elimination to examine whether the main analysis had captured all important associations. Both duration of immune suppression ≤200 cells/μL and ≤100 cells/μL could not both be included together in the model as they were collinear so only ≤200 cells/μL was chosen as I felt in the cART era it was the more clinically relevant cut-off. Current age was included in preference to age at seroconversion as it was more strongly associated with MI incidence, is a well-known predictor of MI and the two variables could not both be included (see Section 4.4.2.3). The only covariate of interest to be retained in the final model was duration of HIV infection which increased MI rates by approximately 70% for each additional decade of infection.

Table 7.16: Sensitivity analysis 6: The association between duration/markers of HIV infection/other factors and MI incidence: mfp combined with backwards elimination (final cycle) [#]

Variable	Model	Vs.	Deviance	Deviance difference	p	Powers	Vs.
Current age~	Null	FP2	2255	46	<0.0001		-2 2
	Linear		2211	3	0.4	1	
	Final		2211			1	
Duration of HIV infection~	Linear	FP2	2211	4	0.3	1	-2 1
	Final		2211			1	
Log_{10} HIV viral load	Null	FP2	2211	7	0.1		-0.5 -0.5
(per log_{10}copy/mL increase*)	Final		2211				
Current CD4 cell count	Null	FP2	2211	7	0.2		-2 -2
(cells/µL*)	Final		2211				
Nadir CD4 cell count	Null	FP2	2211	1	0.8		3 3
(cells/µL*)	Final		2211				
Immune suppression¥	Null	FP2	2211	0.7	1.0		-2 -2
(years ≤200 cells/µL*)	Final		2211				
Years on IDV[¥]	Null	FP2	2211	2	0.8		-2 -2
	Final		2211				
Years on LOP[¥]	Null	FP2	2211	0.5	1.0		-2 -2
	Final		2211				

[#]The following binary and categorical variables were also assessed for inclusion: prior AIDS, prior HCV, calendar period, cohort, apparent ethnicity, mode, sex and current abacavir *All units are those for the linear form of the variable ~per additional 10 years [¥]per additional year

Table 7.17: Sensitivity analysis 6: The associations between duration/markers of HIV infection and MI incidence: mfp combined with backwards elimination (final model)

Variable	IRR (95%CI)	p
Duration of infection~	1.68 (1.20-2.36)	0.002
Current age~	1.87 (1.57-2.22)	<0.0001
Sex		
Male	1	
Female	0.57 (0.31-1.03)	0.07
Cohort		
FRENCH[1] /UKR[2]	1	
AHIVCOS[3]	3.29 (1.40-7.71)	<0.0001
NOR[4]	5.98 (3.44-10.4)	
SEROCCO/SAL[5]	2.89 (1.38-6.06)	
Prior AIDS		
No	1	
Yes	1.72 (1.07-2.77)	0.03
Calendar Period		
≤2000	1	0.1
>2000	2.25 (0.79-0.6.37)	

~Per additional 10 years [1]FHDH/ICONA/PRIMO/CoRIS [2]UKR-UK Register of HIV Seroconverters [3] AHIVCOS-Austrian HIV cohort [4] NOR-Oslo and Ulleval hospital cohort [5]SAL-Southern Alberta Clinic

Sensitivity analysis 7 was restricted to those cohorts (Austrian HIV cohort Study, ICONA, FHDH) which had case definitions and well-validated cases. There were 51 MIs and 12 910 individuals included in the analysis, so 52 events and 5 666 individuals from the main analysis were excluded. Total PYFU was approximately 68% of that of the main analysis (87 007 years compared to 128 654) and the crude first MI rate was 0.6 (0.4-0.8) per 1000 PYFU (vs. 0.8 (0.7-1.0) in the main analysis).

No covariates of interest, apart from duration of HIV infection were associated with MI in the univariable analysis. Unlike the main analysis, nadir CD4 cell count (0.96; 0.84-1.10), duration of immunosuppression at ≤200 cells/μL (1.11; 0.94-1.31) and ≤100 cells/μL (1.17; 0.89-1.55) were not statistically significantly ($p<0.1$) associated with MI. All p values were larger and IRRs were closer to one than in the main analysis due to the reduced power. It was not possible to include duration of lopinavir exposure in the multivariable models for sensitivity analysis 7, because no individuals exposed to lopinavir experienced an event.

In the multivariable analyses shown in Table 7.18, duration of HIV infection (model B_7) was still statistically significantly associated with MI incidence as was prior AIDS (model I_7). The IRR for duration of immune suppression ≤200 cells/μL (model F_7) was the same as in the main analysis, but due to reduced power it was no longer statistically significant. No other covariates of interest were statistically significantly associated with MI.

Table 7.18: Sensitivity analysis 7: Each covariate of interest added separately to multivariable model A (models B_7-I_7) after restricting the MI analysis to those cohorts with case definitions and well-validated cases

Sensitivity model	Covariate	Main analysis IRR (95%CI)*	p	Sensitivity model 7 IRR (95%CI)*	p
B_7	**Duration of HIV Infection~**	2.02 (1.44-2.83)	<0.0001	1.89 (1.16-3.09)	0.01
C_7	**Current CD4 cell count (per 100 cell/µL increase)**	1.00 (0.99-1.02)	0.3	1.01 (0.99-1.02)	>0.1
D_7	**Log viral load (per log_{10} copies/mL increase)**	1.06 (0.92-1.22)	0.4	1.02 (0.82-1.25)	0.9
E_7	**Nadir CD4 cell count (per 100 cell/µL increase)**	0.92 (0.47-1.37)	0.2	1.01 (0.87-1.16)	0.9
F_7	**Duration of immune suppression¥ (≤200cells/µL)**	1.09 (1.01-1.19)	0.04	1.09 (0.92-1.29)	0.3
G_7	**Duration of immune suppression¥ (≤100cells/µL)**	1.12 (0.97-1.32)	0.1	1.17 (0.89-1.53)	0.3
H_7	**Duration of immune suppression¥ (≤50cells/µL)**	1.05 (0.75-1.47)	0.8	0.71 (0.18-2.82)	0.6
I_7	**Prior AIDS (yes vs no)**	2.10 (1.31-3.36)	0.002	2.92 (1.52-5.61)	0.001

*Model A comprised: current age; sex; cohort; indinavir exposure; current abacavir exposure and calendar period ~per additional 10 years ¥per additional year

In the final model, just as in the main analysis (model J) there was evidence that both duration of HIV infection (1.69; 1.03-2.78; p=0.04) and prior AIDS (2.54; 1.30-4.94; p=0.006) were associated with MI incidence after adjustment when included together.

7.7 Summary of my findings

Findings are summarised in Table 7.19. There was evidence of a linear independent association between duration of HIV infection from a well-estimated time of seroconversion and MI incidence. This association was of a similar magnitude to the association with ageing after adjustment. This finding was robust to all sensitivity analyses. This association was independent of HIV viral load and CD4 cell count or any measure derived from CD4. Individuals who had had a prior AIDS defining event also experienced higher rates of MI, including under all sensitivity scenarios.

In the univariable analysis there was evidence that nadir CD4 cell count and duration of immune suppression ≤200/100 cells/μL were associated with MI incidence. These associations were attenuated and not statistically significant however after adjustment. There was no evidence that current CD4 cell count or HIV viral load were associated with MI incidence, but prior AIDS increased incidence.

Table 7.19: Summary of MI analysis findings: The association between markers of HIV infection and MI incidence before and after adjustment and during sensitivity analysis

Exposure of Interest*	Univariable	Adjusted~	Final Multivariable#	Sensitivity Analysis	Comments
Duration of HIV infection (per additional 10 years)	**3.51 (2.59-4.77) p<0.0001**	**2.02 (1.44-2.83) p<0.0001**	**1.87 (1.32-2.64) p<0.0001**	**All SA results similar in size and direction to the main analysis**	Univariable association partially attenuated after adjustment for current age
Nadir CD4 cell count (per 100 cell increase)	**0.82 (0.73-0.92) p=0.001**	0.92 (0.47-1.37) p=0.2	-		Univariable association attenuated after adjustment for current age and cohort
Current CD4 cell count (per 100 cell increase)	1.00 (0.99-1.02) p=0.4	1.0 (0.99-1.02) 2.0 p=0.3	-		
Time spent ≤200 cells/µL (per additional year)	**1.18 (1.10-1.27) p<0.0001**	**1.09 (1.01-1.19) p=0.04**	-		Retained after partial adjustment, but omitted from the final multivariable model as attenuated by inclusion of duration of HIV infection
Time spent ≤100 cells/µL (per additional year)	**1.22 (1.06-1.42) p=0.006**	1.12 (0.97-1.32) p=0.1	-		
Time spent ≤50 cells/µL (per additional 10 years)	1.16 (0.85-1.59) p=0.4	1.05 (0.75-1.47 p=0.8	-		
Current HIV viral Load (per log_{10} cps/mL increase)	0.89 (0.77-1.03) p=0.1	1.06 (0.92-1.22) p=0.4	-		
Prior AIDS diagnosis	**2.92 (1.85-4.61) p<0.0001**	**2.10 (1.31-3.36) p=0.002**	**1.72 (1.06-2.78) p=0.03**		Evidence of an independent association with elevated MI incidence

*All variables were time-updated ~ Each adjusted model (models B-I from Table 7.7) was only adjusted for the following potential confounders: current age, sex, cohort, assumed ethnicity, ART and calendar period. #Adjustment in the final multivariable model (J) which included duration of HIV infection and prior AIDS and was adjusted for all potential confounders previously included in the adjusted model

7.8 A comparison of my study to other studies

7.8.1 Introduction

Table 7.20a and 7.20b compare features of relevant studies including my own. I have already discussed many concepts relating to bias with respect to the fractures analysis. The same mechanisms are likely to be operant across my outcomes. To avoid repetition they will not be discussed again for each either MI or DM, unless features are unique.

There is a small amount of overlap of patients between publications. It is likely that some individuals included in the FHDH case-control study [199] were included in mine. Approximately 10% of my MI analysis population was from ICONA. About 35% of ICONA patients are enrolled in D:A:D. In addition, 2% of individuals in my study were from AHIVCOS (Austrian) and a small number of these are enrolled in D:A:D (via EuroSIDA). VACS-VC and Kaiser Permanente (KP) Northern California (but not KP Southern Californian) contribute to NA-ACCORD [313]. Some overlap between the KP and NA-ACCORD studies is therefore also possible.

7.8.2 Study design and population

All studies with the exception of FHDH (nested case-control) used a cohort design. Most studies used data routinely collected by clinics and hospitals. My study and D:A:D included patients from multiple sites in multiple countries [211]. Other studies included multiple sites in a single country [121, 199, 207, 220] (France & USA) or individuals receiving treatment in a single city [30] (Boston, USA).

7.8.3 Time-interval bias

If HIV is causally associated with MI then it is likely to be mediated through the development of atherosclerosis as a result of chronic ineffective upregulation of inflammatory pathways and lipid dysregulation [325]. There may therefore be substantial time-lag between exposure and outcome. Average follow-up varied across studies from 3-7 years.

7.8.4 Case definitions, misclassification and under-reporting

Table 7.20b summarises MI case definitions and methods of case identification across studies. MI diagnosis is challenging. Clinical signs alone are not a reliable means of identification [285, 327]. A universal (international) AMI definition was formulated in 2000 by an international panel of cardiologists [328]. This definition has been revised (currently

Version 4) as more sensitive tests have become available [327]. Misclassification is therefore potentially a greater source of bias for any MI analysis than it is for my other outcomes (DM and fractures).

Cases were identified in my study, KP, Boston and (~40% of) VACS by finding relevant ICD codes or an MI diagnosis in patient medical records. Insurance database identification of MI using ICD-9 codes (as the KP analysis did) was found in one study to have reasonable sensitivity and specificity (≥86%) [329]. D:A:D defined cases using the old MONICA (WHO) definition which lower sensitivity/specificity than newer definitions [329-331] (see Section 2.1.6). D:A:D used detailed CRFs and end-point review however [332]. The D:A:D MI definition [211, 215] has now been superseded by the revised MONICA WHO (Category A) definition [333]. This additionally requires a record of a marked rise and fall in cardiac biomarkers (preferably troponin) [334]. It is now comparable with the current Universal definition [333]. FHDH applied the American Heart Association case definition (2003) for epidemiological research [254]. Again, based on the Universal definition but definitive cases were not restricted to documented troponin (or other biomarker) rise and fall. Evolving ECG changes were also categorised as definitive [254]. NA-ACCORD and ~60% of VACS MI outcomes were identified using the Universal definition and end-point review. The use of different MI case definitions with varying test sensitivity could contribute to differences in findings [335]. Smoking is a major cause of acute MI [336, 337]. Tiny amounts of heart damage are now detected which previously would have been missed [327, 338].

Only NA-ACCORD restricted the outcome to T1MI [313]. T1MI (atherosclerotic) and T2MI (oxygen supply/demand mismatch) have different risk factors [339]. It seems plausible that associations between my exposures of interest and MI might vary by MI type. In the USA, Crane and colleagues found that nearly half the acute MIs they identified were T2MI. This is far higher than the general population [285]. In their HIV-positive population a third of individuals were PWID in contrast to only ~6% in mine however. Drug use is one of the most important causes of T2MI in those with HIV [285, 339]. If the proportion of T1 and T2MI vary across studies that did not distinguish them, this might explain the conflicting results.The work by Crane et al. also found that just 44% [IQR, 38-49] of adjudicated MI were correctly identified when diagnostic codes were used alone (low sensitivity). These codes (usually ICD) are commonly used to identify cases in studies I have reported, especially those in the US [285].

7.8.5 Confounding

Table 7.20b highlights differences between studies with respect to adjustment for potential confounders. There was considerable variation in: which variables were captured, which adjusted for, how accurate and complete confounder data were, whether completeness of data varied by outcome occurrence, whether adjustment was time-updated or time-fixed at baseline (and baseline differed between studies) and how well all these variations were reported. NA-ACCORD and KP had the most extensive adjustment for important predictors of MI. KP adjusted for socio-economic status (SES) which is known to be a predictor of MI [213]. It is not well-established for most of these factors which are associated with HIV-related variables.

7.8.6 Statistical analysis

FHDH and VACS (2013) undertook MI of relevant missing values [199, 207]. The KP study imputed ethnicity only. All other studies were complete case analyses. I examined categorical, binary and continuous variables and included them based on AIC scores. The VACS (2013) study used established cut-offs for all variables [207]. With the exception of age, which it treated as a continuous, all factors were examined as categories [207]. The VACS (2016) study generated categories which had similar numbers of DM events [221]. Most studies used backwards-stepwise selection during multivariable model building. The VACS (2013) study was the only one to statistically account for the competing risk of death from other causes [207].

All studies reported undertaking sensitivity analyses. FHDH undertook a complete cases analysis and results were similar to the main analysis [199].VACS (2013) [207] undertook sub-group analysis to explore patterns of association across sub-groups. They also explored including less well validated cases (identified through ICD-9 codes) and performed a competing risks analysis [207]. NA-ACCORD reran the analysis omitting those exposed to PIs, which did not change the findings [220]. The Boston Hospital study stratified their analysis by sex [30] to see if associations differed in men and women. D:A:D explored the inclusion/exclusion of outcomes classified as false-positives or "possible" cases by end-point review and also explored the association between CMV and stroke [211].

Table 7.20a: A comparison of characteristics and analyses features of studies examining the association between HIV-related factors and MI

	My study	Kaiser Permanente (KP) [220]	NA-ACCORD [121]	VACS [207]	VACS [221]	Boston Hospitals [203]	FHDH [199]	D:A:D [211]	D:A:D [202]
STUDY CHARACTERISTICS									
Type	Collaboration	Cohort	Collaboration	Cohort		Cohort	Cohort	Cohort	
Country/region	Mainly EU	USA	N. America	USA		USA	France	Mainly EU	
Cohort type(s)	Clinical & Interval (9 cohorts)	Clinical (2 insurance databases)	Clinical & Interval (7 cohorts)	Clinical (8 sites)		Clinical (2 sites)	Clinical (70 sites)	Clinical & Interval (11 cohorts)	
Analysis design	Cohort	Cohort	Cohort	Cohort		Cohort	Case-control	Cohort	
FU period	1988-2013	1996-2009	1995-2014	2003-2010	1996-2012	1998-2008	2000-2006*	1999-2011	1999-2005
Average PYFU‡	5.5 (2.4-10.7)	*4.5†*	3.2 (1.3-5.9)	5.9†	*6.6†*	-	Case-control	*6.7†*	4.5†
Study overlap	FHDH/D:A:D	NA-ACCORD	KP/VACS	NA-ACCORD	NA-ACCORD	None	My study	My study	My study
Publication year	n/a	2014	2017	2013	2016	2010	2012	2013	2007
Demography									
Age: FU start	33 [28-40]	40-44#	40-49#	48†	46†	46 (SD:12)	46 [40-54]	38†	39 (34-45)
Sex: Male	80%	91%	86%	97%	97%	69%	89%	74%	76%
Ethnicity: Black	10%	18%	36%	48%	55%	24%	-	-	17%
Annual loss to FU	4%								<3%
DATA SOURCES									
Outcome	(E)MR	EMR	EMR	EMR/registry		EMR	(E)MR	CRFs	
Linkage	-	Insurance database/ Registry	Insurance databases/Registries			Insurance database	-	EMR/ Registry	

‡ In italics if a mean otherwise median *Case-control study-cases were captured between these dates # Reported in age bands SD-standard deviation EMR-Electronic medical records
† No 95%CIs/IQRs reported ≠

Table 7.20b: A comparison of characteristics and analyses features of studies examining the association between HIV-related factors and MI

OUTCOMES	My study	Kaiser (KP) Permanente[220]	NA-ACCORD [121]	VACS [207]	VACS [221]	Boston Hospitals [203]	FHDH [199]	D:A:D [202, 211]
Means of ID	Recorded in (E)MR	Abstracted from EMR	Abstracted from EMR	EMR/death registry/ admin. database	Abstracted from EMR	Abstracted insurance data	Abstracted from EMR	Detailed CRF
MI case definition	MI in notes	ICD-9 Code 410.x (Acute MI)	T1MI Universal definition	Death cert./Universal definition/ICD-9/	ICD-9 code 410.xx	ICD-9 code 410.xx	Universal definition	WHO MONICA
Validation/end-point review	No	No	Yes	61% of MI validated & reviewed≠	Yes/No‡	No	Yes	Yes
Crude rate /1000 PYFU	0.8 (0.7-1.0)	**2.8 (2.5-3.2)**	2.6 (2.3-2.9)#	-	-	-	-	3.2 (3.0-3.4)
Exposures of interest	A-E	A-D	B-D	B,D	B,D	B-D	C,D	B,C,E
POTENTIAL CONFOUNDERS/STATISTICS*								
Not included/captured	1-10	6-9	2, 8-10	8-10	6-15	2, 8-11	6, 9-11,13	6, 9,10
Time-updated	0,11-15,16	0,3-5,12,16	0,1,3-6	--	16	-	-	1,2
Time-fixed	-	1, 2, 13-15	11-14†,16	0-6,11,13-15	0,1,3-5	0,1,3-6,12-14	0-5,7,8,12,14-16	0,3-5,7,8
Missing data reported	All	For ethnicity only	Not reported	All	3,6¥	Not reported	All	1,2
Data missing for >10%	11	-	-	3	3,6	-	3, 8	1,2
Missing data handling	Complete case	Complete case/ MI for ethnicity	Complete case	MI	Complete case	Complete case	MI (& sensitivity analysis)	Complete case
Method of variable selection	Some *a priori* + stepwise (<0.1)	*A priori*	*A priori*	*A priori* for relevant model	AIC score + size/strength	*A priori*	Combination: *a priori* & stepwise	*A priori*
Sensitivity analysis	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

UDMI-Universal definition of myocardial infarction ICD-9 International Classification of Diseases-9 CRF-case report form MI-Multiple imputation (E)MR- (electronic) medical records PPV-positive predictive value NPV-negative predictive value #After type-2 MI had been removed (~50%) ≠ A sensitivity analysis including only validated cases was also undertaken ¥-Not included in multivariable model however ‡Positive predictive value 82/Negative predictive value 100 *Confounders accounted for numbered as follows: 0-age 1-smoking, 2-BMI, 3-Lipids/cholesterol, 4-BP, 5-DM, 6-CKD/eGFR, 7-prior CVD, 8-parental CVD, 9-atrial fibrillation, 10-SES/deprivation 11-HCV/HBV 12-ART 13-ethnicity 14-sex 15-alcohol/drug abuse 16-calendar period †ART and HCV omitted from final model A-duration of infection, B-current CD4, C-Nadir CD4, D-Current Viral Load, E-Duration of immunosuppression

7.8.7 A comparison of findings across studies

Tables 7.21a-e summarise results for each of my exposures of interest. There was generally a lack of agreement in findings across studies. Only two studies (mine and KP) examined the association between duration of HIV infection and MI [220] (Table 7.21a). Results were conflicting. Neither my study nor D:A:D found an association between duration of immune-suppression and MI. All studies found that a higher nadir CD4 cell count was associated with a small decrease in the incidence/odds of an MI after adjustment, but some studies lacked power so the findings were not statistically significant (Table 7.21b).

Table 7.21a: A summary of findings: The association between duration of HIV infection and MI before and after adjustment

			Duration of HIV infection				Factors accounted for in multivariable analysis*																							
Study (Location)	**MI type**	**N/n HIV positive /events**	**Point Estimate**	**Increments in years**	**Point Estimate (95% CI) p-value**		**Age**	**Sex**	**Mode**	**Ethnicity**	**Smoking**	**BMI**	**Lipids**	**Hypertension**	**Diabetes**	**Family history [a]**	**Prior**	**Alcohol**	**Cocaine**	**eGFR**	**Medication [b]**	**Co-morbidities [c]**	**HCV/HBV**	**ART**	**Nadir CD4 count**	**HIV viral load**	**Current CD4 count**	**Prior AIDS**
					Univariable	**Multivariable**																							
My study (EU/ Canada)	All	18 576 103	IRR	Per 10	**3.51 (2.59-4.77) <0.0001**	**1.87 (1.32-2.64) p=<0.0001**	**√**	√	√	√	x	x	x	x	x	x	x	x	x	x	x	x	**√**	**√**	**√**	**√**	**√**	**√**
Kaiser Permanente (US) [220]	Acute (T1&2)	22 081 280	IRR	≥10 5.0-<10 <5	≠	0.92 (0.67-1.27) 0.6 1.05 (0.77-1.45) 0.8 1.00 [¥≠]	**√**	√	√	√	√	√	**√**	**√**	**√**	x	x	√ [d]	√ [d]	x	x	x	x	**√** [e]	√	**√**	**√**	x

*By restriction, matching, adjustment, randomisation or found not to be associated with the outcome and then dropped**. √** Time-updated variables have bold ticks ≠ Values not reported [¥] Duration of infection was calculated as time since first positive test HCV-Hepatitis C Virus HBV-Hepatitis B Virus [a] CVD in parent <60 years of age, [b] Medication: steroids, antipsychotics & anti-hypertensives [c] Comorbidities: severe mental illness, systemic lupus, erectile dysfunction, rheumatoid arthritis, migraines & atrial fibrillation [d] drug and alcohol abuse were combined [e] only adjusted for prior ART (yes vs no)

Table 7.21b: A summary of findings: The association between nadir CD4 cell count and MI before and after adjustment

		Nadir CD4 cell count					Factors accounted for in multivariable analysis*																								
Study (Location)	**MI type**	**N/n (HIV positive/ events)**	**Point Estimate**	**Unit† Increase Cells/µL**	**Point Estimate (95% CI) p-value**		**Age**	**Sex**	**Mode**	**Ethnicity**	**Smoking**	**BMI**	**Lipids**	**Hypertension**	**Diabetes**	**Family history [a]**	**Prior CVD/CVE**	**Alcohol**	**Cocaine**	**eGFR**	**Medication [b]**	**Co-morbidities [c]**	**HCV/HBV**	**ART**	**HIV infection duration**	**HIV viral load**	**Current CD4 cell count**	**Prior AIDS**
					Univariable	**Multivariable**																									
My study (EU/ Canada)	All	18 576 103	IRR	100	**0.82 (0.73-0.92)** **0.001**	0.92 (0.47-1.37) 0.2	**√**	√	√	√	x	x	x	x	x	x	x	x	x	x	x	x	**√**	**√**	**√**	**√**	**√**	**√**
KP (US) [220]	Acute	22 081 280	IRR	100	≠	**0.88 (0.81-0.96)** **0.006**	**√**	√	√	√	√	√	**√**	**√**	**√**	x	x	√ [d]	√ [d]	x	x	x	x	√ [e]	x	**√**	**√**	x
Boston Hospitals (US) [203]	Acute	6 517 273	OR	50	≠	0.95 (0.89-1.01) 0.09	√	√	x	√	√	x	√	√	√	x	√	x	x	√	x	√	x	**√**	√	**√**	**√**	x
FHDH (France) [199]	Acute	1 173 289	OR	Per log_2 increase	≠	**0.90 (0.83-0.97)≠**	**√**	**√**	√	**√**	**√**	√	**√**	**√**	**√**	√	**√**	**√**	**√** [d]	**√**	**√**	**√**	x	√	x	**√**	**√**	**√**
D:A:D [211] (Majority EU)	Acute	33301 716	IRR	<100 100-199 200-299 300-399 400-499 ≥500	≠	1.15 (0.93, 1.43) 0.96 (0.76, 1.21) 1.0 (Reference) 0.79 (0.57, 1.09) 0.62 (0.38, 1.01) 1.11 (0.75, 1.63)	√	√	√	√	√	√	√	√	√	√	√	**x**	**x**	**x**	**x**	**x**	x	**√**	x	**x**	**x**	**x**
D:A:D [202]	Acute	23 437 345	IRR	50	0.98 (0.95-1.01)	-[f]																							

*By restriction, matching, adjustment, randomisation or found not to be associated with the outcome and then dropped**. √** Time-updated variables with bold ticks ≠ Values not reported HCV-Hepatitis C Virus HBV-Hepatitis B Virus [a] CVD in parent <60 years of age, [b] Medication: steroids, antipsychotics & anti-hypertensives [c] Comorbidities: severe mental illness, systemic lupus, erectile dysfunction, rheumatoid arthritis, migraines & atrial fibrillation [d] drug and alcohol abuse were combined [e] only adjusted for prior ART (yes vs no) at baseline [f] Value after adjustment not reported

Table 7.21c: A summary of findings: The association between current CD4 cell count and MI before and after adjustment

			Current CD4 cell count				Factors accounted for in multivariable analysis*																				
Study (Location)	**MI type**	**N/n (HIV positive/ events)**	**Point Estimate**	**Unit Increase Cells/µL†**	**Point Estimate (95% CI) p-value**		**Age**	**Sex**	**Mode**	**Ethnicity**	**Smoking**	**BMI**	**Lipids**	**Hypertension**	**Diabetes**	**Family MI [a]**	**Prior CVD**	**Alcohol**	**Cocaine**	**eGFR**	**Medication [b]**	**Comorbidities[c]**	**HCV/HBV**	**ART**	**HIV duration**	**Viral load**	**Nadir CD4**	**Prior AIDS**
					Univariable	**Multivariable**																						
My study (EU/ Canada)	All	18 576 103	IRR	100	1.00 (0.99-1.02) 0.4	1.00 (0.99-1.02) 0.3	**√**	√	√	√	x	x	x	x	x	x	x	x	x	x	x	x	**√**	**√**	**√**	**√**	**√**	**√**
KP (US) [220]	Acute	22 081 280	IRR	100	≠	1.03 (0.97-1.10) 0.3	**√**	√	√	√	√	√	**√**	**√**	**√**	x	x	√ [d]	√ [d]	x	x	x	x	**√** [e]	x	**√**	**√**	x
NA-ACCORD (US) [121]	Acute (T1 only)	29,169 335	IRR	<100 100-199 200-349 350-499 ≥500	≠	**2.19 (1.44-3.33)** **1.60 (1.09, 2.34)** **1.37 (1.01, 1.86)** 1.32 (0.98, 1.77) 1.00≠	**√**	√	√	√	√	x	**√**	**√**	**√**	x	x	x	x	**√**	x	√	√	**√**	**√**	x	**√**	**√**
VACS (US) [221]	Acute	8 168 196	HR	<200 200-349 350-499 ≥500	**1.68 (1.14, 2.49)[f]** 1.37 (0.93, 2.01)[f] 1.39 (0.95, 2.04)[f] vs 1.00[f≠]	**1.58 (1.06-2.35)** 1.39 (0.95-2.04) 1.31 (0.89-1.93) 1.00≠	√	x [g]	x	x	√	x	√	√	√	x	√	x	x	x	x	x	x	√ [h]	x	√	x	x
VACS [207]	Acute	27 350 363		<200 vs.≥200	≠	**1.57 (1.10-2.24)≠**	√	√	x	√	√	√	√	√	x	x	√	√	√	√	x	x	√	√ [j]	x	x	x	x
Boston Hospitals (US) [203]	Acute	6 517 273	OR	<200 vs.≥200	**2.00 (1.48-2.71)** **<0.0001**	**1.74 (1.07-2.81)** **0.02**	√	√	x	√	√	x	√	√	√	x	√	x	x	√	x	√	x	**√**	√	**√**	**√**	x
D:A:D [211]	Acute	33301 716	IRR	100 100-199 200-299 300-399 400-499 ≥500	≠	1.07 (0.71-1.64) 0.98 (0.71-1.36) 1.00 (Ref.)≠ 1.03 (0.79-1.34) 0.88 (0.66-1.16) 0.90 (0.71-1.14)	√	√	√	√	√	√	√	√	√	√	√	**x**	**x**	**x**	**x**	**x**	x	**√**	x	**x**	**x**	**x**

*By any means**. √** Time-updated ≠ Not reported HCV-Hepatitis C Virus HBV-Hepatitis B Virus [a] CVD in parent <60 years, [b] Medication: steroids, antipsychotics & anti-hypertensives [c] Comorbidities: mental illness, lupus, erectile dysfunction, rheumatoid arthritis, migraines & atrial fibrillation [d] drug and alcohol combined [e] only adjusted for prior ART (yes vs no) [f] adjusted for age only [g] not adjusted for but sample was 97% male [h] restricted to on cART

Table 7.21d: A summary of findings: The association between current HIV viral load and MI before and after adjustment

			HIV viral load				Factors accounted for in multivariable analysis*																				
Study (Location)	**MI type**	**HIV positive /events (N/n)**	**Point estimate**	**Per log_{10} copies/mL increase†**	**Point Estimate (95% CI) p-value**		**Age**	**Sex**	**Mode**	**Ethnicity**	**Smoking**	**BMI**	**Lipids**	**Hypertension**	**Diabetes**	**Family history [a]**	**Prior CVD/CVE**	**Alcohol**	**Cocaine**	**eGFR**	**Medication [b]**	**Co-morbidities [c]**	**HCV/HBV**	**ART**	**HIV infection**	**Nadir CD4**	**Current CD4**	**Prior AIDS**
					Univariable	**Multivariable**																						
My study (EU/ Canada)	All	18 576 103	IRR	100	0.89 (0.77-1.03) 0.1	1.06 (0.92-1.22) 0.4[d]	**√**	√	√	√	x	x	x	x	x	x	x	x	x	x	x	x	**√**	**√**	**√**	**√**	**√**	**√**
Kaiser Permanente (US) [220]	Acute	22 081 280	IRR	100	≠	1.03 (0.97-1.08) 0.4	**√**	√	√	√	√	√	**√**	**√**	**√**	x	x	√ [d]	√ [d]	x	x	x	x	**√** [e]	x	**√**	**√**	x
NA-ACCORD (US) [121]	Acute (T1 only)	29,169 335	IRR	≥400 vs. <400	≠	1.20 (0.92-1.56)≠ **(1.36 (1.06–1.75)†**	**√**	√	√	√	√	x	**√**	**√**	**√**	x	x	x	x	**√**	x	√	√	**√**	**√**	x	**√**	**√**
VACS (US) [221]	Acute	8 168 196	HR	≤200 201-999 1000-9999 ≥100,000	1≠ **1.74 (1.09, 2.79)** 1.15 (0.66, 2.01) 1.34 (0.88, 2.04)	1≠ **1.71 (1.06-2.74)** 1.11 (0.64-1.93) 1.30 (0.85-1.99)	√	x [g]	x	x	√	x	√	√	√	x	√	x	x	x	x	x	x	√ [h]	x	√	√	x
VACS [207]	Acute	27 350 508	HR	≥500 vs. <500	≠	**1.60 (1.14-2.22)**	√	√	x	√	√	√	√	√	x	x	√	√	√	√	x	x	√	√ [i]	x	x	x	x
Boston Hospitals (US) [203]		6 517 273	OR	>100,000 vs ≤100,000	**2.23 (1.37-3.65)** **0.001**	1.63 (0.91-2.93) 0.1	√	√	x	√	√	x	√	√	√	x	√	x	x	√	x	√	x	**√**	√	**√**	**√**	x
FHDH (France) [199]	Acute	1 173 289	OR	>50 vs ≤50	**1.40 (1.07-1.83)** **0.06‡**	**1.51 (1.09-2.10)≠**	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	x	√	x	√	√	√

*By restriction, matching, adjustment, randomisation or found not to be associated with the outcome and then dropped**. √** Time-updated variables with bold ticks ≠ Values not reported † With current CD4 cell count omitted from the adjustment ‡ p value reported, OR not reported, but calculated from data provided HCV-Hepatitis C Virus HBV-Hepatitis B Virus [a] CVD in parent <60 years of age, [b] Medication: steroids, antipsychotics & anti-hypertensives [c] Comorbidities: severe mental illness, systemic lupus, erectile dysfunction, rheumatoid arthritis, migraines & atrial fibrillation [d] Only adjusted for potential confounders (not for other covariates of interest) [e] only adjusted for prior ART (yes vs no) [g] not adjusted for but sample was 97% male [h] restricted to those on cART [i] adjusted for PI exposure only (but baseline ART class/regimen not found to be associated)

Table 7.21e: A summary of findings: The association between duration of immunosuppression and MI

			Duration of immune-suppression				Factors accounted for in multivariable analysis*																								
Study (Location)	**MI type**	**N/n (HIV positive/ events)**	**Point Estimate**	**Time spent ≤200 cells/ µL**	**Point Estimate (95% CI) p-value**		**Age**	**Sex**	**Mode**	**Ethnicity**	**Smoking**	**BMI**	**Lipids**	**Hypertension**	**Diabetes**	**Family history [a]**	**Prior CVD/CVE**	**Alcohol**	**Cocaine**	**eGFR**	**Medication [b]**	**Co-morbidities [c]**	**HCV/HBV**	**ART**	**HIV infection duration**	**HIV viral load**	**Current CD4**	**Nadir CD4**	**Prior AIDS**
					Univariable	**Multivariable**																									
My study (EU/ Canada)	All	18 576 103	IRR	Per additional year	**1.18 (1.10-1.27) <0.0001**	1.00 (0.99-1.02) 0.3 [d]	**√**	√	√	√	x	x	x	x	x	x	x	x	x	x	x	x	**√**	**√**	**√**	**√**	**√**	**√**	**√**
D:A:D (Mainly EU) [211]	Acute (T1&2)	33 301 716	IRR	≤2	≠	1.09 (0.91, 1.31)	√	√	√	√	√	√	√	√	√	√	√	x	x	x	x	x	x	**√**	x	x	x	x	x
				>2-≤4		**1.33 (1.04, 1.70)**																								
				>4-≤6		1.33 (0.98, 1.81)																								
				>6-≤8		0.86 (0.54, 1.35)																								
				>8-≤10		**1.61 (1.01, 2.57)**																								
				>10		1.13 (0.61, 2.13)																								

*By restriction, matching, adjustment, randomisation or found not to be associated with the outcome and then dropped. **√** Time-updated variables have bold ticks ≠ not reported HCV-Hepatitis C Virus HBV-Hepatitis B Virus [a] CVD in parent <60 years of age, [b] Medication: steroids, antipsychotics & anti-hypertensives [c] Comorbidities: severe mental illness, systemic lupus, erectile dysfunction, rheumatoid arthritis, migraines & atrial fibrillation [d] Adjusted for potential confounders, AIDS and duration of infection

8 Chapter 8: Diabetes Mellitus Incidence

8.1 Introduction

Diabetes mellitus is a common disease, both in the general population and in those with HIV [232, 340]. It causes significant morbidity and increases the probability of an early death [341]. It is unclear whether higher rates of DM occur in those with HIV [123, 124]. It has been purported that ART, HCV co-infection or the sequelae of HIV infection itself may increase the risk of DM in those with HIV [134, 228, 342]. This chapter explores the association between HIV specific factors and DM incidence before and after adjustment.

I present the results of my analysis to examine the associations between HIV-related factors and DM incidence. In Section 8.2 I report which individuals are included in the analysis and provide numbers and reasons for exclusion. I then go on to describe baseline characteristics (section 8.3) before presenting my main results (section 8.4 & 8.5). Section 8.6 shows the results of my sensitivity analyses and is followed by a summary of my findings (Section 8.7). I then go on to compare my study characteristics and findings with those of other studies (Section 8.8).

8.2 Inclusion

In total 9 012 HIV-positive individuals with 109 first DM diagnoses from nine cohorts were potentially eligible for inclusion in my analysis. Figure 8.1 provides details of individuals/events included.

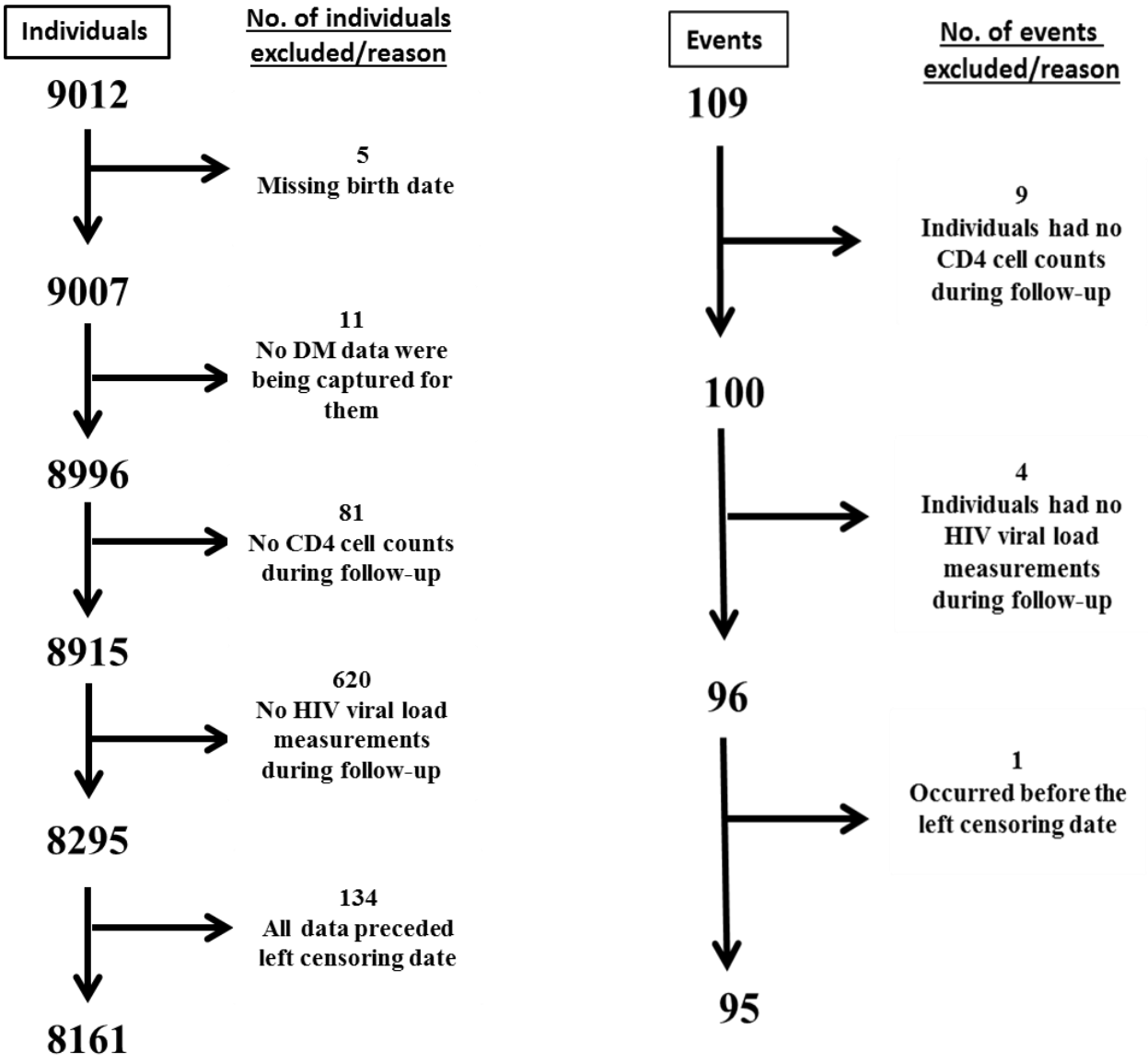


Figure 8.1: HIV-positive individuals and events included in my DM analysis with numbers and reasons for exclusions

In total 14 events (13%) and 851 (9%) individuals were excluded from the analysis. There were 8 161 individuals included with a total follow-up time of 55 270 person years (median

5.4 [2.3-10.9] years). There were 95 first DM diagnoses (event rate, per 1000 PYFU, 1.93; 95%CI, 1.58-2.34).

8.3 Baseline characteristics

Table 8.1 presents baseline characteristics of individuals (HIV seroconversion) both overall and by sex. The median age at seroconversion was 31.6 [26.4-38.7] years with males being slightly older than females. A higher percentage of females were of black ethnicity (14.8%) than males (4.8%). Transmission in men was predominantly MSM (80.7%) and in females was largely through heterosexual sex (76.6%). The UK Register (32.4%) and ICONA (22.8%) contributed the largest number of subjects.

Table 8.1: Baseline characteristics of individuals in an analysis examining the association between duration/markers of HIV infection and DM incidence, stratified by sex

	Male (N=7 055)	Female (N=1 111)	Total (N=8 166)
	Median [IQR] or n (%≠)		
Age at seroconversion (years)	32.1 (26.7-39.0)	29.4 (24.2-36.0)	31.6 (26.4-38.7)
Ethnicity			
Black	240 (3.7)	152 (15.0)	392 (5.3)
White	5 979 (93.4)	842 (82.9)	6 821 (91.9)
Other	185 (2.9)	22 (2.2)	207 (2.8)
Missing*	651	95	746
Assumed ethnicity^			
Black	328 (4.8)	157 (14.8)	485 (6.1)
White	6 309 (92.3)	878 (82.9)	7 187 (91.0)
Other	200 (2.9)	24 (2.3)	224 (2.8)
Missing	218	52	270
Mode			
MSM#/Bisexual female	5 562 (80.7)	2 (0.2)	5 564 (69.8)
PWID~	422 (6.1)	215 (19.8)	637 (8.0)
Heterosexual	713 (10.3)	832 (76.6)	1 545 (19.4)
Other	193 (2.8)	37 (3.4)	230 (2.9)
Missing	165	25	190
Cohort			
AMACS	308 (4.4)	12 (1.1)	320 (3.92)
AHIVCOS	315 (4.5)	82 (7.4)	397 (4.9)
ICONA	1 460 (20.7)	400 (36.0)	1 860 (22.8)
NOR	432 (6.1)	69 (6.2)	501 (6.1)
PRIMO	1 142 (16.2)	200 (18.0)	1 342 (16.4)
SEROCO	368 (5.2)	105 (9.5)	473 (5.8)
SAL	183 (2.6)	26 (2.3)	209 (2.6)
UKR	2 461 (34.9)	187 (16.8)	2 648 (32.4)
CoRIS	386 (5.5)	30 (2.7)	416 (5.1)

*Missing ethnicity also included those for whom the collection of ethnicity data were prohibited. ^Region of origin was used as a proxy for ethnicity where ethnicity was missing. AMACS-Athens Multicentre AIDS Cohort Study AHIVCOS- Austrian HIV cohort ICONA-Italian Cohort of Antiretroviral Naïve Patients NOR-Oslo and Ulleval hospital cohort PRIMO-Primary Infection Cohort ANRS CO6 SEROCO-Seroconverter cohort ANRSCO2 SAL-Southern Alberta Clinic UKR-UK Register of HIV Seroconverters CoRIS-The Cohort of the Spanish HIV Research Network #MSM, men who have sex with men ~PWID intravenous drug users ≠Percentage of non-missing total

8.4 Univariable analysis

Univariable analysis included an exploration of both time-fixed and time-updated factors and examined what form best described them (categorical, continuous, binary and log transformed).

8.4.1 Time-fixed variables

In the univariable analysis of time-fixed variables (Table 8.2), DM rates increased with older age at HIV seroconversion. I could not include an <20 year age category as I had done in the other analyses as there was only a single case of DM in this age category. Individuals ≥60 years at seroconversion experienced nearly seven times the rate of DM when compared to individuals aged <40 years, although confidence intervals were wide due to the small number of individuals in the ≥60 years category. No statistical evidence was found for an association between sex, ethnicity or mode of HIV acquisition and DM. Rates were similar across cohorts, with the exception of Southern Alberta (SAL) and the Austrian cohort (AHIVCOS), whose patients had a nearly three-fold higher incidence rate when compared to those in the UK register.

Table 8.2: The association between time-fixed variables and DM incidence in HIV-positive individuals: Rates and univariable analysis

Time-fixed factor	Event (n)	PYFU*	Rate/1000 PYFU (95%CI)	IRR (95%CI)	p[1]
Seroconversion age (years)					
<40	53	44 748	1.18 (0.90-1.55)	1	<0.0001
40-<60	37	9 910	3.73 (2.71-5.15)	3.15 (2.07-4.80)	
≥60	5	612	8.17 (3.40-19.6)	6.90 (2.76-17.3)	
Sex					
Male	85	46 342	1.83 (1.48-2.27)	1	0.1
Female	10	8 930	1.12 (0.60-2.08)	0.61 (0.32-1.17)	
Ethnicity					
White	85	48 229	1.76 (1.42-2.18)	1	0.6
Black	2	2 374	0.84 (0.21-3.37)	0.48 (0.12-1.94)	
Other	1	1 065	0.93 (0.13-6.67)	0.53 (0.07-3.83)	
Missing	7	3 603	1.94 (0.93-4.08)	1.10 (0.51-2.38)	
Likely ethnicity[2]					
White	85	49 139	1.73 (1.40-2.14)	1	0.4
Black	2	2 586	0.77 (0.19-3.09)	0.45 (0.11-1.82)	
Other	1	1 154	0.87 (0.12-6.15)	0.50 (0.07-3.60)	
Missing	7	2 391	2.93 (1.40-6.14)	1.69 (0.78-3.66)	
Mode					
MSM[#]	55	35 249	1.56 (1.20-2.03)	1	0.6
PWID~	11	4 764	2.31 (1.28-4.16)	1.48 (0.77-2.82)	
Heterosexual	22	11 224	1.96 (1.29-2.98)	1.26 (0.77-2.06)	
Other	5	2 929	1.71 (0.71-4.10)	1.09 (0.44-2.73)	
Missing	2	1 105	1.81 (0.45-7.24)	1.16 (0.28-4.76)	
Cohort					
UKR/CoRIS	28	21 120	1.32 (0.92-1.92)	1	0.09
AMACS	6	3 164	1.90 (0.85-4.22)	1.43 (0.59-3.45)	
AHIVCOS	12	3 035	3.95 (2.25-6.96)	2.98 (1.52-5.87)	
ICONA	17	9 862	1.72 (1.07-2.77)	1.30 (0.71-2.37)	
NOR	7	4 324	1.62 (0.77-3.40)	1.22 (0.53-2.80)	
PRIMO	12	7 396	1.62 (0.92-2.86)	1.22 (0.62-2.41)	
SEROCO	9	5 271	1.71 (0.88-3.28)	1.29 (0.61-2.73)	
SAL	4	1 098	3.64 (1.37-9.70)	2.75 (0.96-7.83)	

[#]MSM-Men who have sex with men ~PWID-Intravenous drug users [1]Test for heterogeneity excluding missing [2] Region of origin was used as a proxy for ethnicity when ethnicity was missing. UKR-UK register of HIV seroconverters AMACS-Athens Multicentre AIDS Cohort Study AHIVCOS- Austrian HIV cohort ICONA-Italian Cohort of Antiretroviral Naïve Patients NOR-Oslo and Ulleval hospital cohort PRIMO-Primary Infection Cohort ANRS CO6 SEROCO-Seroconverter cohort ANRSCO2NOR-Oslo and Ulleval hospital cohort SAL-Southern Alberta Clinic *PYFU-person-years of follow-up

8.4.2 Time-updated variables

Duration of infection was associated with DM in the univariable analysis. DM rates were over three times higher in those infected >12 years previously when compared to those infected ≤ 3 years ago (Table 8.3, first page of 3 page table). Current age was also strongly positively associated with DM in a dose-response manner. Those >60 years of age experienced over 17 times the incidence when compared to those ≤40 years.

There was no evidence of an association between current CD4 cell count as a categorical variable and DM incidence, but those with nadir CD4 cell counts ≤100cells/μL had an ~80% increased DM incidence when compared to those with nadirs >100 cells/μL (p=0.02).

There was evidence of an association between HIV viral load category and DM in the univariable analysis (p=0.04). Those with HIV viral load <1000 copies/mL, (expected to be those on ART), experienced the highest rates.

Both current exposure to ART and duration of ART exposure (p=0.02 and p<0.0001 respectively) were associated with increased rates of DM in the univariable analysis. Those with >9 years of ART experienced a 3.5-fold increase in rates compared to those never exposed to ART. Both current exposure and duration of exposure to PI or NRTI (considered separately) were associated with increased rates of DM. The largest effect size was seen for duration of PI exposure, with those with >4 years on PIs experiencing over three times the rate of DM when compared to those never exposed. No association was found between current NNRTI exposure and DM rates. Those with NNRTI exposure duration of >4 years experienced approximately double the rate of DM (weak evidence, p=0.07). Current and cumulative exposure to AZT increased DM incidence. There was no statistical evidence that current exposure to d4T was associated with DM rates, although there was limited exposure (patients accrued only 1 512 PYFU on d4T of 55 270 (2.7%) in total) so statistical power was limited. There was weak evidence (p=0.09) that cumulative exposure to d4T was associated with increased DM incidence in the univariable analysis. Those with >1 year of prior d4T exposure experienced a 74% increase in incidence, although confidence intervals were wide. No individuals experienced an event whilst on current ddI, but longer duration of historical exposure to ddI was associated with increased DM rates (Table 8.3, 3rd page). Those with >1 year of ddI exposure experienced more than 2.5 times the rate of DM than those never exposed.

DM rates increased with longer duration of immunosuppression ≤200 cells/µL as a categorical variable, with >12 months duration being associated with an approximate doubling in the incidence in univariable analysis (p=0.02).

Evidence for an association between prior AIDS and DM was lacking (p=0.3). Previous HCV infection was found to be associated with an 80% increase in DM incidence.

Current follow-up in the >2000 calendar period was associated with increased DM incidence (which may represent improved data capture on DM status by cohorts).

Table 8.3 (Page 1 of 3): The association between time-updated factors and DM incidence in HIV-positive individuals: Rates and univariable analysis

Time-updated factor	Event No.	PYFU*	Rate/1000 PYFU* (95%CI)	IRR (95%CI)	p[1]
Duration of infection					
≤3 years	14	14 032	1.00 (0.59-1.68)	1	0.0008
>3-6	16	11 863	1.35 (0.83-2.20)	1.35 (0.66-2.77)	
>6-9	11	9 280	1.19 (0.66-2.14)	1.19 (0.54-2.62)	
>9-12	15	7 314	2.05 (1.24-3.40)	2.06 (0.99-4.26)	
>12	39	12 781	3.05 (2.23-4.18)	3.06 (1.66-5.63)	
Current age					
≤40 years	15	28 965	0.52 (0.31-0.86)	1	<0.0001
>40-50	34	17 854	1.90 (1.36-2.67)	3.68 (2.00-6.75)	
>50-60	28	6 467	4.33 (2.99-6.27)	8.36 (4.47-15.65)	
>60	18	1 985	9.07 (5.71-14.4)	17.45 (8.80-34.6)	
CD4 cell count (cells/µL)					
>500	56	28 818	1.94 (1.50-2.52)	1	0.7
351-500	21	13 314	1.58 (1.03-2.42)	0.81 (0.49-1.34)	
201-350	12	8 591	1.40 (0.79-2.46)	0.72 (0.39-1.34)	
101-200	5	2 292	2.18 (0.91-5.24)	1.12 (0.45-2.80)	
≤100	1	1 159	0.86 (0.12-6.12)	0.42 (0.06-3.21)	
Missing	0	1 095	0.00	-	
Nadir CD4 cell count (cells/µL)					
≤100	19	6 613	2.87 (1.83-4.50)	1	0.08
101-200	17	10 198	1.67 (1.04-2.68)	0.58 (0.30-1.12)	
201-350	33	18 005	1.83 (1.30-2.58)	0.64 (0.36-1.12)	
351-500	10	10 573	0.95 (0.51-1.76)	0.33 (0.15-0.71)	
>500	16	8 787	1.82 (1.11-2.97)	0.63 (0.33-1.23)	
Missing	0	1 095	0.00	-	
Binary nadir CD4 count (cells/µL)					
>100	76	47 563	1.60 (1.28-2.00)	1	0.02
≤100	19	6 613	2.87 (1.83-4.50)	1.83 (1.11-3.04)	
Missing	0	1 095	-	-	
HIV viral load (copies/mL)					
<1000	69	31 370	2.20 (1.74-2.78)	1	0.04
1000-9999	5	7 259	0.69 (0.29-1.65)	0.31 (0.13-0.78)	
10000-99999	14	10 319	1.36 (0.80-2.29)	0.62 (0.35-1.10)	
≥100000	7	3 468	2.02 (0.96-4.23)	0.92 (0.42-2.00)	
Missing	0	2 855	0	-	

[1] Test for heterogeneity [2] ART-Antiretroviral therapy *Person-years of follow-up

Table 8.3 (continued page 2 of 3): The association between time-updated factors and DM incidence in HIV-positive individuals: Rates and univariable analysis

Factor	Event No.	PYFU*	Rate/1000(95%CI)	IRR (95%CI)	p[1]
Current ART[2]					
No	22	19 017	1.16 (0.76-1.76)	1	0.02
Yes	73	36 253	2.01 (1.60-2.53)	1.74 (1.08-2.80)	
Years on ART[2]					
No ART	17	16 402	1.04 (0.64-1.67)	1	<0.0001
≤3 years	16	15 011	1.07 (0.65-1.74)	1.03 (0.52-2.04)	
>3-6 years	14	9 476	1.48 (0.87-2.49)	1.43 (0.70-2.89)	
>6-9 years	19	6 515	2.92 (1.86-4.57)	2.82 (1.46-5.42)	
>9 years	29	7 867	3.69 (2.56-5.30)	3.56 (1.96-6.47)	
Current PI					
No	53	38184	1.39 (1.06-1.82)	1	0.006
Yes	42	17087	2.46 (1.82-3.33)	1.77 (1.18-2.66)	
Years on PI					
Never exposed	35	31062	1.13 (0.81-1.57)	1	<0.0001
≤2	17	9872	1.72 (1.07-2.77)	1.53 (0.86-2.73)	
>2-4	9	4580	1.97 (1.02-3.78)	1.74 (0.84-3.63)	
>4	34	9756	3.49 (2.49-4.88)	3.09 (1.93-4.96)	
Current NRTI					
No	31	22974	1.35 (0.95-1.92)	1	0.08
Yes	64	32296	1.98 (1.55-2.53)	1.47 (0.96-2.26)	
Years of NRTI					
Never exposed	22	18425	1.19 (0.79-1.81)	1	0.0004
≤2	8	11374	0.70 (0.35-1.41)	0.59 (0.24-1.01)	
>2-4	19	7483	2.54 (1.62-3.98)	2.13 (1.15-3.93)	
>4	46	17995	2.56 (1.91-3.41)	2.14 (1.29-3.56)	
Current NNRTI					
No	72	40447	1.78 (1.41-2.24)	1	0.6
Yes	23	14823	1.55 (1.03-2.34)	0.87 (0.55-1.39)	
Years of NNRTI					
Never exposed	50	34325	1.46 (1.10-1.92)	1	0.07
≤2	16	8876	1.80 (1.10-2.94)	1.24 (0.70-2.17)	
>2-4	7	4352	1.61 (0.77-3.37)	1.10 (0.50-2.44)	
>4	22	7717	2.85 (1.88-4.33)	1.96 (1.19-3.23)	
Current AZT					
No	91	50829	1.79 (1.46-2.20)	1	<0.0001
Yes	4	4441	0.90 (0.34-2.40)	1.25 (1.24-1.26)	
Years of AZT					
Never exposed	61	42428	1.44 (1.12-1.85)	1	0.03
≤0.5	13	3514	3.70 (2.15-6.37)	2.57 (1.41-4.68)	
>0.5-1	7	1718	4.07 (1.94-8.55)	2.83 (1.30-6.20)	
>1	14	7610	1.84 (1.09-3.10)	1.28 (0.72-2.29)	
Current d4T					
No	93	53758	1.73 (1.41-2.12)	1	0.7
Yes	2	1512	1.32 (0.33-5.29)	0.76 (0.19-3.10)	
Years of d4T exposure					
Never exposed	71	46007	1.54 (1.22-1.95)	1	0.09
≤ 1	11	4420	2.49 (1.38-4.49)	1.61 (0.85-3.04)	
>1	13	4843	2.68 (1.56-4.62)	1.74 (0.96-3.14)	

[1] Test for heterogeneity * Person-years of follow-up

Table 8.3 (Page 3 of 3): The association between time-updated factors and DM incidence in HIV-positive individuals: Rates and univariable analysis

Time-updated factor	Event No.	PYFU*	Rate/1000 (95%CI)	IRR (95%CI)	p[1]
Current ddI					
No	95	52 890	1.80 (1.47-2.20)	1	-
Yes	0	2 381	-	-	
Years of ddI exposure					
Never exposed	57	42 728	1.33 (1.03-1.73)	1	0.03
≤ 1	15	5 883	2.55 (1.54-4.23)	1.91 (1.08-3.38)	
>1	23	6 659	3.45 (2.30-5.20)	2.59 (1.60-4.20)	
Duration of immune suppression ≤200 cells/µL					
Always >200	60	37 430	1.60 (1.24-2.06)	1	0.02
>0-6 months	8	7 198	1.11 (0.56-2.22)	0.69 (0.33-1.45)	
>6-12 months	5	2 625	1.90 (0.79-4.58)	1.19 (0.48-2.96)	
>12 months	22	6 922	3.18 (2.09-4.83)	1.98 (1.22-3.23)	
Missing	0	1 095			
Duration of immune suppression ≤100 cells/µL					
Always >100	76	47 588	1.60 (1.27-2.00)	1	0.2
>0-6 months	9	3 011	2.99 (1.56-5.74)	1.87 (0.94-3.73)	
>6-12 months	3	1 019	2.95 (0.95-9.13)	1.84 (0.58-5.85)	
>12 months	7	2 558	2.74 (1.30-5.74)	1.71 (0.79-3.72)	
Missing	0	1 095	-	-	
Duration of immune suppression ≤50 cells/µL					
Always >50	87	50 430	1.73 (1.40-2.13)	1	0.5
>0-6 months	4	1 854	2.16 (0.81-5.75)	1.25 (0.46-3.41)	
>6 months	4	1 891	2.12 (0.79-5.64)	1.23 (0.45-3.34)	
Missing	0	1 095	-	-	
Prior AIDS					
No	83	50059	1.66 (1.34-2.06)	1	0.3
Yes	12	5211	2.30 (1.31-4.06)	1.39 (0.76-2.54)	
Prior HCV					
No	69	45697	1.51 (1.19-1.91)	1	
Yes	26	9573	2.72 (1.85-3.99)	1.80 (1.15-2.82)	0.01
Calendar Period					
≤2000	4	7400	0.54 (0.20-1.44)	1	0.03
>2000-2005	26	14700	1.77 (1.20-2.60)	3.27 (1.14-9.38)	
>2005-2010	44	19068	2.31 (1.72-3.10)	4.27 (1.53-11.9)	
>2010	21	14108	1.49 (0.97-2.28)	2.75 (0.95-8.02)	
Binary Calendar Period					
≤2000	4	7399	0.54 (0.20-1.44)	1	
>2000	91	47871	1.90 (1.55-2.33)	3.52 (1.29-9.57)	0.01

[1]Test for heterogeneity * Person-years of follow-up

There was evidence in univariable analysis ($p<0.0001$) that DM incidence was associated with increasing duration of HIV infection, age at seroconversion and current age (all as continuous variables). With each additional decade of both duration of HIV infection and current age the incidence of DM more than doubled.

There was little evidence that current or nadir CD4 cell count was associated with DM incidence.

There was an inverse association between DM rates and $\log_{10}$ HIV viral load. This association was attenuated and no longer statistically significant when current age was added to the model (IRR, 1.02; 95%CI, 0.88-1.18; p=0.8).

Duration of immunosuppression ≤200 cells/µL was associated with DM incidence. Rates were 10% higher for each additional year without adjustment (p=0.02).

With the exception of AZT and d4T, there was evidence that all measures of duration of exposure to ART (ART, PI, NRTI, NNRTI and ddI) were associated with increased DM incidence in univariable analysis.

Although there was evidence of increasing fracture rates with increasing calendar year, the association was found not to be linear as it plateaued in later calendar years. The binary variable (≤2000 vs. >2000) was used in subsequent multivariable analysis (AIC 2063 for binary vs. 2064 for linear).

Table 8.4: The association between continuous variables and DM incidence in HIV-positive individuals: univariable analysis*

	IRR (95%CI)	p
Duration of HIV infection~	2.14 (1.56-2.92)	<0.0001
Seroconversion age~	1.86 (1.58-2.19)	<0.0001
Current age~	2.53 (2.09-3.06)	<0.0001
Current CD4 cell count (per 100 cell/μL increase)	1.02 (0.95-1.10)	0.5
Nadir CD4 cell count (per 100 cell/μL increase)	0.94 (0.84-1.05)	0.3
Current Log HIV viral load (per log_{10} copies/mL increase)	0.86 (0.74-1.00)	0.05
Years of immune-suppression¥ ≤200 cells/μL	1.10 (1.02-1.19)	0.02
Years of immune-suppression¥ ≤100 cells/μL	1.12 (0.94-1.33)	0.2
Years of immune-suppression¥ ≤50 cells/mL	1.18 (0.88-1.57)	0.3
Years of ART¥	1.11 (1.07-1.16)	<0.0001
Years of PI¥	1.13 (1.08-1.17)	<0.0001
Years of NRTI¥	1.12 (1.07-1.16)	<0.0001
Years of NNRTI¥	1.07 (1.02-1.13)	0.01
Years of AZT¥	0.99 (0.88-1.11)	0.9
Years of d4T¥	1.04 (0.86-1.26)	0.7
Years of ddi¥	1.23 (1.11-1.37)	<0.0001
Calendar year¥	1.04 (1.00-1.09)	0.04

*All variables were time-updated apart from age at seroconversion ~per additional 10 years ¥per additional year

Both age at seroconversion and current age were strongly positively correlated (Pearson's correlation coefficient, r_P =0.8) and could not be included in the same model. Age at seroconversion was only weakly negatively correlated with duration of HIV infection (r_P=0.2). Current age was only weakly positively correlated with duration of HIV infection (r_P=0.3).

During exploration of association and correlation between independent variables I found that duration of NRTI exposure was positively correlated with both duration of HIV infection (r_P =0.7) and duration of PI exposure (r_P =0.6). After adjusting for current age the effect size was slightly larger and strength of the association was stronger for duration of PI exposure. Therefore duration of PI was included in multivariable model A0 and duration of NRTI exposure was omitted due to collinearity.

Ethnicity was not found to be associated with DM either in univariable or multivariable analysis and so it was omitted.

8.5 Multivariable analysis

Table 8.5 shows the basic multivariable model A0. It includes demographic factors (age and sex), potential confounders determined *a priori* where data were available (years of PI, HCV) and both cohort and calendar period. Factors found to be associated with DM (p <0.05) were: current age, cohort, duration of PI and prior HCV. Weak evidence of an association was observed between sex (p=0.09) and DM. I found no evidence that calendar period was associated with DM after adjustment (p=0.2).

Table 8.5: The association between demographic factors/potential confounders and DM incidence in HIV-positive individuals: Basic multivariable model A0

Variable	IRR (95%CI)	p
Current age~	2.44 (1.99-2.99)	<0.0001
Sex		
Male	1	
Female	0.57 (0.29-1.10)	0.09
Cohort		
UKR/CoRIS	1	
All other cohorts	1.63 (1.04-2.57)	0.03
Duration of PI exposure¥	1.07 (1.02-1.12)	0.02
Prior HCV infection		
No	1	
Yes	1.91 (1.20-3.03)	0.007
Calendar period		
≤2000	1	0.2
>2000-2005	2.04 (0.73-5.70)	

~per additional 10 years ¥ per additional year UKR- UK Register All other cohorts comprised: AMACS, ICONA, PRIMO, SEROCO, SAL and AHIVCOS †Region of origin was used as a proxy for ethnicity where ethnicity was missing

Interaction was then assessed between each variable included in model A0 and current age and sex; the interaction term p values are shown in Table 8.6. There was no statistical evidence for interaction.

As there was no evidence of interaction, the final baseline model (model A) was identical to model A0 (shown in Table 8.5).

Table 8.6: The assessment of interaction between age and sex and other variables in the basic multivariable model A0
(which examines the association between demographic factors/potential confounders and DM incidence in HIV-positive individuals)

Variable	Interaction term p values*	
	Current age	Sex
Sex	0.4	X
Cohort	0.5	0.4
Years on PI	0.7	0.4
Prior HCV infection	0.2	0.7
Calendar period	0.9	0.4

*Interaction terms were added individually to model A0 and tested for significance X- not applicable

Table 8.7: The association between duration/markers of HIV infection and DM incidence in PLHIV: IRRs for each covariate of interest added separately to multivariable model A (Model B-J) with univariable results shown for comparison

Model*	Covariate	Multivariable IRR (95%CI)*	p	Univariable IRR (95%CI)	p
B	**Duration of HIV Infection~**	1.02 (0.68-1.52)	0.9	2.14 (1.56-2.92)	<0.0001
C	**Current CD4 cell count (per 100 cell/µL increase)**	1.01 (0.94-1.09)	0.7	1.00 (0.93-1.07)	0.9
D	**Log viral load (per log copies/mL increase)**	1.09 (0.93-1.26)	0.3	0.86 (0.74-1.00)	0.05
E	**Nadir CD4 cell count (per 100 cell/µL increase)**	1.07 (0.96-1.21)	0.2	0.92 (0.83-1.02)	0.1
F	**Years of immune suppression¥ (≤200cells/µL)**	1.01 (0.92-1.10)	0.9	1.10 (1.01-1.19)	0.02
G	**Years of immune suppression¥ (≤100cells/µL)**	1.00 (0.82-1.21)	1.0	1.12 (0.94-1.33)	0.2
H	**Years of immune suppression¥ (≤50cells/µL)**	1.04 (0.74-1.45)	0.8	1.18 (0.88-1.57)	0.3
I	**Prior AIDS (yes vs no)**	0.94 (0.51-1.75)	0.8	1.39 (0.76-2.54)	0.3
J	**Binary Nadir CD4 cell count (≤100 cell/ µL vs >100)**	1.20 (0.73-1.98)	0.5	1.83 (1.11-3.04)	0.02

~per additional 10 years ¥per additional year * Each model was adjusted for: current age, sex, cohort, years on PI, HCV-seropositivity and calendar period ¥ per additional year

No exposures of interest were found to be significant predictors of DM incidence after adjustment. Duration of infection was forced into the final model however, as it was the primary exposure of interest (Table 8.8).

The final step in the main analysis was to add all previously excluded covariates back into the analysis one at a time to ensure none met the criteria for re-inclusion. None did. Previously unassessed interactions were then evaluated, but no interactions (p<0.1) were found.

Table 8.8: The association between duration/markers of HIV infection and DM incidence in HIV-positive individuals after adjusting for demographic factors/potential confounders: The final multivariable model (model J)

Variable	IRR (95%CI)	p
Duration of HIV infection~	1.02 (0.68-1.52)	0.9
Current age~	2.43 (1.98-3.00)	<0.0001
Sex		
Male	1	
Female	0.57 (0.29-1.10)	0.09
Cohort		
UKR/CoRIS	1	
All Other cohorts	1.64 (1.04-2.59)	0.04
Duration of PI exposure¥	1.07 (1.02-1.12)	0.01
Prior HCV		
No	1	
Yes	1.89 (1.16-3.08)	0.01
Calendar period		
≤2000	1	
>2000	2.04 (0.73-5.70)	0.2

~per additional 10 years UKR-UK Register All other cohorts comprised: AMACS, ICONA, PRIMO, SEROCO, SAL and AHIVCOS

8.6 Sensitivity analyses

Sensitivity analysis 1a explored the effects of censoring follow-up after the first gap in CD4 measurements of at least a year (Table 8.9).

Table 8.9: Sensitivity analysis 1a: Each CD4 related covariate added separately (model C/E-H) to model A after censoring follow-up after the first gap in CD4 measurements of ≥1 year (with main results for comparison)

Sensitivity model	Covariate	Main analysis IRR (95%CI)*	p	Sensitivity model 1_a IRR (95%CI)*	p
C_{1a}	**Current CD4 cell count (per 100 cell/µL increase)**	1.01 (0.94-1.09)	0.7	1.00 (0.99-1.01)	0.4
E_{1a}	**Nadir CD4 cell count (per 100 cell increase)**	1.07 (0.96-1.21)	0.2	1.02 (0.93-1.17)	0.7
F_{1a}	**Years of immune suppression¥ (≤200cells/µL)**	1.01 (0.92-1.10)	0.9	**1.11 (1.01-1.23)**	**0.04**
G_{1a}	**Years of immune suppression¥ (≤100cells/µL)**	1.00 (0.82-1.21)	1.0	**1.24 (1.04-1.48)**	**0.02**
H_{1a}	**Years of immune suppression¥ (≤50cells/µL)**	1.04 (0.74-1.45)	0.8	**1.43 (1.07-1.91)**	**0.02**

*Each model also included: current age, sex, cohort, ethnicity, years on PI, years, HCV-seropositivity, calendar period ¥per additional year

Results were similar to those of the main analysis (Table 8.7) for current and nadir CD4. However, there was evidence that cumulative exposure to immune-suppression at all three cut-offs (≤200/100/50 cells/µL) was associated with DM incidence after censoring when CD4 measurements became a year out of date.

Sensitivity analysis 1b explored the effects of censoring follow-up after the first gap in CD4 measurements of at least a year, but allowing follow-up to re-start if subsequent CD4 measurements were taken (Table 8.10).

Table 8.10: Sensitivity analysis 1b: Each CD4 related covariate added separately (models C/E-H) to model A after censoring follow-up after the first gap in CD4 measurements of ≥1 year, but allowing re-entry to follow-up if subsequent measurements were recorded (with main analysis results for comparison)

Sensitivity model	Covariate	Main analysis IRR (95%CI)*	p	Sensitivity model 1_b IRR (95%CI)*	p
C_{1b}	**Current CD4 cell count (per 100 cell/µL increase)**	1.01 (0.94-1.09)	0.7	1.00 (0.99-1.01)	0.9
E_{1b}	**Nadir CD4 cell count (per 100 cell increase)**	1.07 (0.96-1.21)	0.2	1.07 (0.97-1.19)	0.2
F_{1b}	**Years of immune suppression¥ (≤200cells/µL)**	1.01 (0.92-1.10)	0.9	1.02 (0.92-1.13)	0.7
G_{1b}	**Years of immune suppression¥ (≤100cells/µL)**	1.00 (0.82-1.21)	1.0	1.06 (0.87-1.28)	0.6
H_{1b}	**Years of immune suppression¥ (≤50cells/µL)**	1.04 (0.74-1.45)	0.8	1.14 (0.81-1.59)	0.4

*Each model was also adjusted for: current age, sex, cohort, years on PI, HCV-seropositivity, and calendar period ¥per additional year

Results for sensitivity analysis 1b were similar with respect to size, direction and strength of statistical evidence to those of the main analysis (Table 8.7) for all covariates of interest examined.

Sensitivity analysis 2a explored the effects of censoring follow-up after the first gap in HIV viral load measurements of at least a year (Table 8.11).

Table 8.11: Sensitivity analysis 2a: HIV viral load added to model A after censoring follow-up after the first gap in HIV viral load measurements of ≥1 year (with main analysis results for comparison)

Sensitivity model	Covariate	Main analysis IRR (95%CI)*	p	Sensitivity model 2_a IRR (95%CI)*	p
D_{2a}	Log_{10} HIV viral load (per log_{10} copies/mL increase)	1.09 (0.93-1.26)	0.3	1.10 (0.93-1.30)	0.3

The model was adjusted for: current age, sex, cohort, years on PI, HCV-seropositivity and calendar period

The results of sensitivity analysis 2a were very similar to those of the main analysis. Sensitivity analysis 2b explored the effects of censoring follow-up after the first gap in HIV viral load measurements of at least a year, but allowing follow-up to re-start if subsequent viral load measurements were taken (Table 8.12).

Table 8.12: Sensitivity analysis 2b: HIV viral load added to model A after censoring follow-up after the first gap in HIV viral load measurements of ≥1 year but allowing re-entry to follow-up if subsequent measurements were recorded (with main analysis results for comparison)

Sensitivity model	Covariate	Main analysis IRR (95%CI)*	p	Sensitivity model 2_b IRR (95%CI)*	p
D_{2b}	Log_{10} HIV viral load (per log_{10} copies/mL increase)	1.09 (0.93-1.26)	0.3	1.11 (0.95-1.30)	0.2

The model was adjusted for: current age, sex, cohort, years on PI, HCV-seropositivity and calendar period

Again, the results of sensitivity analysis 2b were similar to those of the main analysis with respect to size, direction and the strength of statistical evidence.

A sensitivity analysis was also undertaken where individuals did not enter follow-up until 01/01/2005 (although exposure status was accrued since seroconversion) when clinicians became more aware of the importance of collecting data on SNAEs such as DM (Table 8.13).

Table 8.13: Sensitivity analysis 3: Each covariate of interest added separately to multivariable model A (models B3-I3) after restricting follow-up to ≥01/01/2005 (with main analysis results for comparison)

Model*	Covariate	Multivariable IRR (95%CI)*	p	Sensitivity model 3 IRR (95%CI)	p
B_3	Duration of HIV Infection~	1.02 (0.68-1.52)	0.9	1.03 (0.63-1.69)	0.9
C_3	Current CD4 cell count (per 100 cell/µL increase)	1.01 (0.94-1.09)	0.7	1.03 (0.93-1.13)	0.6
D_3	Log viral load (per log copies/mL increase)	1.09 (0.93-1.26)	0.3	0.92 (0.73-1.17)	0.5
E_3	Nadir CD4 cell count (per 100 cell/µL increase)	1.07 (0.96-1.21)	0.2	1.05 (0.90-1.22)	0.5
F_3	Years of immune suppression¥ (≤200cells/µL)	1.01 (0.92-1.10)	0.9	1.02 (0.91-1.14)	0.8
G_3	Years of immune suppression¥ (≤100cells/µL)	1.00 (0.82-1.21)	1.0	1.01 (0.79-1.30)	0.9
H_3	Years of immune suppression¥ (≤50cells/µL)	1.04 (0.74-1.45)	0.8	1.11 (0.73-1.67)	0.6
I_3	Prior AIDS (yes vs no)	0.94 (0.51-1.75)	0.8	1.21 (0.58 2.52)	0.6

*Each model was also adjusted for: current age, sex, cohort, years on PI, HCV-seropositivity, and calendar period ¥per additional year

The association with duration of infection remained similar in the multivariable analysis adjusting for all variables in Model A (IRR, 1.03; 95%CI, 0.63-1.69; p=0.9). All other covariates of interest also had associations of a similar magnitude and in the same direction as in the main analysis and none were statistically significantly associated with DM.

Due to some uncertainty as to when systematic data collection began for some cohorts, a sensitivity analysis was undertaken to explore changing the assumptions about left-censoring dates. In sensitivity analysis 4 all cohorts were left-censored the day after the first recorded DM event for that cohort (Table 8.14).

Table 8.14: Sensitivity analysis 4: Each covariate of interest added separately to multivariable model A (models B_4-I_4) after restricting entry to follow-up by cohort to the day after the first DM was recorded for that cohort (main analysis results included for comparison)

Sensitivity model	Covariate	Main analysis IRR (95%CI)*	p	Sensitivity model 4 IRR (95%CI)*	p
B_4	**Duration of HIV Infection~**	1.02 (0.68-1.52)	0.9	0.90 (0.58-1.40)	0.7
C_4	**Current CD4 cell count (per 100 cell/µL increase)**	1.01 (0.94-1.09)	0.7	1.02 (0.94-1.10)	0.6
D_4	**Log viral load (per log_{10} copies/mL increase)**	1.09 (0.93-1.26)	0.3	1.10 (0.94-1.30)	0.2
E_4	**Nadir CD4 cell count (per 100 cell/µL increase)**	1.07 (0.96-1.21)	0.2	1.09 (0.96-1.22)	0.2
F_4	**Duration of immune suppression¥ (≤200cells/µL)**	1.01 (0.92-1.10)	0.9	1.00 (0.90-1.11)	0.9
G_4	**Duration of immune suppression¥ (≤100cells/µL)**	1.00 (0.82-1.21)	1.0	0.99 (0.79-1.25)	0.9
H_4	**Duration of immune suppression¥ (≤50cells/µL)**	1.04 (0.74-1.45)	0.8	1.01 (0.66-1.54)	1.0
I_4	**Prior AIDS**	0.94 (0.51-1.75)	0.8	1.06 (0.55-2.03)	0.9

*Each model was also adjusted for: current age, sex, cohort, years on PI, HCV-seropositivity, and calendar period ¥per additional year

In-line with the main analysis, sensitivity analysis 4 found no evidence that duration of HIV infection was associated with DM (IRR, 0.90; 95%CI, 0.58-1.40; p=0.7) after multivariable adjustment for all covariates in Model A. A similar lack of evidence for an association with was found for all other covariates of interest.

The final sensitivity analyses explored the possibility that associations were non-linear. A Multivariable fractional polynomial (mfp) backwards-stepwise regression was undertaken. In sensitivity analysis 5 mfp transformations were applied to the existing variables in the final model (J) of the main analysis (Table 8.15). In sensitivity analysis 6 mfp was combined with an automated backwards-stepwise selection ($p<0.1$) process (Table 8.16 and 8.17).

Table 8.15: Sensitivity analysis 5: Assessing fractional transformations (mfp) of variables included in the final multivariable analysis (model J) on the association between duration/markers of HIV infection and DM*

Variable	Model	vs.	Deviance difference	p	Powers	vs.
Duration of Infection~	Linear	FP2	3.3	0.4	1	-2 -2
	Final				1	
Current age~	Linear	FP2	3.4	0.3	1	-1 -3
	Final				1	
Duration of PI exposure¥	Linear	FP2	1.2	0.7	1	3 3
	Final				1	

Total model deviance=1952 *Current age, sex, cohort, years on PI, HCV-seropositivity, calendar period

In sensitivity analysis 5 no transformations were found to improve the strength of associations between covariates and the DM when compared to the untransformed model. Therefore, the final model for SENSITIVITY ANALYSIS 5 was identical to the final model of the main analysis (model J).

In sensitivity analysis 6 (Tables 8.16 and 8.17) all available variables were included in the backwards stepwise model initially, with a few exceptions. Firstly, age at seroconversion and current age were collinear (r=0.8) and could not be included together so current age was chosen, as this could more fully account for the effect of ageing. The same was true for duration of PI and NRTI exposure and so duration of PI exposure was included. Immunosuppression at 200/100/50 cells/µL were also collinear (100 & 200 r_p=0.8, 50 & 100 r_p=0.9 and 200 & 50 r_p=0.6) and so 200 cell/µL was included as it had the strongest statistical evidence for an association in univariable analysis. Models were re-run with NRTI exposure duration in place of PI exposure duration and the other cut-offs for immune-suppression in place of 200. This was to check that these changes did not alter findings. No changes to findings were found.

Table 8.16: Sensitivity analysis 6: The association between duration/markers of HIV infection/other factors and DM incidence: mfp combined with backwards elimination (final cycle) [#]

Variable	Model	vs.	Deviance difference	p	Powers	vs.
Duration of HIV infection~	Linear	FP2	0.7	0.9	1	0.5 0.5
	Final				1	
Current age~	Null	FP2	80	<0.0001		-2 -2
	Linear		3	0.3	1	
	Final				1	
Duration of PI exposure[¥]	Null	FP2	7	0.2	.	3 3
	Final				.	
HIV viral load (copies/mL)	Null	FP2	5	0.3	.	1 3
	Final				.	
Current CD4 cell count	Null	FP2	2	0.8	.	-0.5 3
(cells/µL)	Final				.	
Nadir CD4 cell count	Null	FP2	4	0.4	.	1 1
(cells/µL)	Final				.	
Immune suppression[¥]	Null	FP2	3	0.5	.	-2 -2
(years ≤200 cells/µL*)	Final				.	

[#]The following binary/categorical variables were also assessed for inclusion: prior AIDS, prior HCV, calendar period, cohort, mode, ethnicity and sex *All units are those for the linear form of the variable ~per additional 10 years [¥]per additional year

The association between all covariates included in the final model and DM were found to be best described by including them in their linear form (untransformed). The final multivariable model (Table 8.17) found no evidence of an association between duration of HIV infection and DM incidence after adjustment. There was evidence however, that increasing current age, male sex, HCV-seropositivity and cohorts other than the UK Register/ CORIS were associated with elevated DM incidence.

Table 8.17: Sensitivity analysis 6: The associations between duration/markers of HIV infection and DM incidence: mfp combined with backwards elimination (final model)

Variable	IRR (95%CI)	p
Duration of infection~	1.19 (0.83-1.69)	0.3
Current age~	2.63 (2.13-3.25)	<0.0001
Sex		
Male	1	
Female	0.47 (0.22-0.97)	0.04
Cohort		
UKR/CORIS	1	
All other	1.71 (1.04-2.81)	0.03
Prior HCV		
No	1	
Yes	2.00 (1.20-3.30)	0.008

~Per additional 10 years. All other cohorts comprise: AMACS, ICONA, PRIMO, SEROCO, Southern Alberta Clinic and AHIVCOS

8.7 Summary of my findings

Table 8.18 summarises my results. DM incidence approximately doubled for every additional 10 years of HIV infection in univariable analysis. The size and strength of this association was markedly attenuated and no longer statistically significant however, after adjusting for current age. These findings were robust to all sensitivity analyses. Similarly, no evidence was found that DM was associated with either current or nadir CD4 cell count.

Duration of immune-suppression at ≤200/100/50 cells/μL was associated with increased DM incidence in the univariable analysis. Severe immune-suppression (<100/50 cells/μL) did not occur commonly, however. In the multivariable analysis and in all sensitivity analyses (with the exception of sensitivity analysis 1a) evidence for an association between DM and immune-suppression duration was not apparent. In sensitivity analysis 1a there was evidence that duration of immune-suppression at all three cut-offs was associated with an increased incidence of DM and that the effect size increased with the severity of immune dysfunction. This sensitivity analysis censored follow-up at the first instance of a gap in CD4 cell count measurements of more than a year. This study found little evidence that there was an association between current log HIV viral load or prior AIDS and DM in any of the multivariable models (main analysis or sensitivity analysis).

Table 8.18: Summary of DM analysis findings: The association between markers of HIV infection and DM incidence before and after adjustment and during sensitivity analysis

Exposure of Interest*	Univariable	Adjusted~	Final Multivariable#	Sensitivity Analysis (SAs)	Comments
Duration of HIV infection (per additional 10 years)	**2.14 (1.56-2.92) p<0.0001**	1.02 (0.68-1.52) p=0.9	1.02 (0.68-1.52) p=0.9	All SA similar size/direction to main multivariable analysis	Univariable association was attenuated by adjustment for both current age and duration of PI exposure
Nadir CD4 cell count (per 100 cell/µL increase)	0.92 (0.83-1.02) p=0.1	1.07 (0.96-1.21) p=0.2	-	All SA similar size/direction to main multivariable analysis	
Current CD4 cell count (per 100 cell increase)	1.00 (0.93-1.07) p=0.9	1.01 (0.94-1.09) p=0.7	-	All SA similar size/direction to main multivariable analysis	
Time spent ≤200 cells/µL (per additional year)	**1.10 (1.02-1.19) p=0.02**	1.01 (0.92-1.10) p=0.9	-	**SA 1a: 1.11 (1.01-1.23) p=0.04**	SA 1a: Follow-up was censored after any gap of >1 year between CD4 measurements. In SA1a the size of the association between duration of immune-suppression (at all thee cut-offs) and DM increased both before and after adjustment (when compared to the main analysis)
Time spent ≤100 cells/µL (per additional year)	1.12 (0.94-1.33) p=0.2	1.00 (0.82-1.21) p=1.0	-	**SA 1a: 1.24 (0.99-1.45) p=0.02**	
Time spent ≤50 cells/µL (per additional year)	1.18 (0.88-1.57) p=0.3	1.04 (0.74-1.45) p=0.8	-	**SA 1a: 1.43 (1.07-1.91) p=0.02**	
Current log HIV viral Load (per log_{10} cps/mL increase)	0.86 (0.74-1.00) p=0.05	1.09 (0.93-1.26) p=0.3		All SA similar size/direction to main multivariable analysis	
Prior AIDS diagnosis (yes vs no)	1.39 (0.76-2.54) p=0.3	0.94 (0.51-1.75) p=0.8	-	All SA similar size/direction to main multivariable analysis	

*All variables were time-updated ~Each adjusted model (models B-I from Table 8.7) was adjusted for the following potential confounders: current age, sex, and cohort, duration of PI exposure and calendar period. #Adjustment in the final multivariable model (J) which included duration of HIV infection and prior AIDS and also adjusted for all potential confounders previously included in the adjusted model

8.8 A comparison of my study to other studies

8.8.1 Introduction

Table 8.19a and 8.19b compare study characteristics and analysis features of studies which examined the association between relevant exposures and DM.

8.8.2 Study design and population

The majority of individuals in my analysis were from: Greece, Austria, Italy, Norway, Spain, UK and France. Canada (Southern Alberta Cohort) also contributed a small percentage of patients (3%). Data from one large hospital in Italy was analysed in the San Raffaele Hospital study [229]. The Medicaid study was confined to South Carolina, USA [232]. Other studies have already been described with respect to fractures or MI.

8.8.3 Sampling (ascertainment) bias

In the analysis undertaken by Petoumenos et al. for D:A:D, only those with a complete DM risk profile were included [226]. This enabled the investigators to compare risk prediction models directly: the primary aim of the analysis. Around half of the cohort was excluded however, which had the potential to lead to strong sampling bias. I compared individuals in this analysis to those included in the D:A:D paper by De Wit et al. which included nearly all D:A:D patients [158, 226]. Those in Petoumenos's paper were more likely to be white (61% vs. 45%) but had a similar sex (27% vs. 26% female) and mode (41% vs. 43% MSM) distribution.

8.8.4 Loss to follow-up & time-interval bias

Loss to follow-up information was only available for my study (~4% annually), MACS (27% over the total follow-up period) [124]and for D:A:D (5-8% annually) [158, 226]. In my analysis 15% of follow-up was ≥10 years. Other studies only provided average follow-up, which was ≥4 years for all studies I examined.

8.8.5 Case definitions, misclassification and under-reporting

8.8.5.1 *Case definitions*

Diabetes is relatively straightforward to diagnose based on clinical signs and blood glucose, however current ways of classifying diabetes into types is imperfect [343]. Type-1 DM occurs due to an absolute lack of insulin as a result of auto-immune destruction of the cells which produce it in the pancreas [344]. It commonly occurs in childhood. Type-2 DM is due to the development of insensitivity to insulin [344]. Type-2 (but not type-1) is strongly linked to being overweight or obese and to genetic factors [134, 225].

My definition relied on a DM diagnosis in the clinical records with a date. I did not distinguish between types [343]. The D:A:D analyses [158, 226] was restricted to type-2 DM and divided type-2 DM cases into confirmed and probable. Confirmed cases required a record of fasting glucose >7 on two separate occasions. A probable case was one reported by clinician in the medical records with a date [158, 226]. The Swiss study also restricted their DM analysis to type-2 and had a very similar definition to the D:A:D as SHCS contributes to D:A:D [224, 225]. Table 8.19b includes a summary of DM case definitions and how cases were detected. The MACS and South Carolina studies included both types of DM [124, 232]. South Carolina included a sensitivity analysis with only type-2 and results were very similar to the main analysis [232]. The San Raffaele study included type-2 only [229].

The lack of distinction between types of DM in some case definitions (including my own) might have contributed a little to the heterogeneity in findings between studies. However type-1 DM does not commonly develop in adulthood and pre-existing DM cases were omitted from all analyses [343, 344]. Over 90% of cases of DM in the general population (including children) are type-2 [134, 343-345].The impact of including type-1 DM on findings is therefore likely to be fairly small.

8.8.5.2 *Misclassification/under-reporting*

Both the proportion of individuals with DM in the general population and the proportion of them undiagnosed (which can be as high as 30%) are highly variable across geographical areas [346-348]. It is possible that diagnosis rates are higher in those with HIV compared to the general population however. This could occur as a results of regular HIV-related clinic visits (commonly at least twice a year) where blood tests are routinely taken in those receiving HIV care in resource rich settings.

Glucose intolerance (unlike fractures or MI) is not a binary phenomenon. Type-2 DM is diagnosed at a standardised blood glucose threshold, but individuals are actually on a spectrum of glucose (in)tolerance [349]. All insulin resistant individuals with elevated blood glucose below the threshold would be classified as not having the outcome. This could reduce the apparent strengths of associations for causal exposures (especially those with a dose-response relationship).

Blood glucose measurements are routinely taken in HIV clinics and so under-reporting of DM outcomes maybe less likely than for MI or fractures.

D:A:D and the SHCS validated DM cases using detailed CRFs thereby reducing the risk of misclassification [158, 224-226] (Table 8.19b). MACS relied on both routine blood tests and regular patient questionnaires to detect DM in its patients [124]. San Raffaele reviewed cases to reduce misclassification risk [229]. South Carolina relied on ICD-9 codes without review which increased the risk of misclassification [232].

8.8.6 Confounding

The factors adjusted for across studies were highly variable (summarised in Table 8.19b). The MACS study only adjusted for time-fixed values at study entry (first blood glucose after 1999) [124].

Risk factors for type-2 DM are well established [227] and include: age, ethnicity, BMI, smoking, deprivation, family history, CVD, hypertension and steroids [227]. More recent risk factors identified which have not commonly been adjusted for include: severe mental illness, antipsychotic medications, learning disability, statins, polycystic ovaries, prior fasting blood glucose values and glycated haemoglobin (HBA1c) value [227]. The latter two values are likely to be on the causal pathway. Which of these factors are associated with my exposures of interest is not well established.

8.8.7 Statistical Analysis

All DM studies undertook a complete case analysis. I reported the percentage of missing data for each variable. Three studies (D:A:D 2008, SHCS 2010 and South Carolina) did not report numbers with missing data for any variables (either for individual factors or overall) [158, 225, 232]. Petoumenos et al. (D:A:D) reported the total number of patients not included in the final multivariable model due to missing data (~50%) [226]. MACS and SHCS (2007) reported levels of missing data for some variables [124, 225].

As with MI and fractures almost all studies (except for my own and the SHCS 2007 study [224]) did not simultaneously report both univariable and multivariable results. This made

it impossible to evaluate the effect of adjustment on findings in most cases. The Italian study reported all univariable results and multivariable results where p≤0.05 however [229]. This enabled me to see where adjustment had changed the results in their study.

The only study to statistically account for the competing risk of death was the San Raffaele study [229].

SHCS 2007 explored collinearity and discussed it with respect to concomitant NRTI and PI use [224]. De Wit et al. discussed collinearity between cholesterol and triglycerides in their D:A:D study, so must have examined it [158]. I examined collinearity between all variables in my study. Other studies did not report examining it, but publication word counts make inclusion of all relevant details very difficult.

A lack of interaction between PI and NRTI was mentioned by Ledergerber et al. so interactions must have been explored in their analyses like my own [224]. No other studies mentioned interaction.

Time-varying confounding was only addressed in the South Carolina Medicaid study by their use of marginal structural models [232]. The association found between PI exposure and DM was reported as very similar using standard Cox regression when compared to the causal model in sensitivity analysis [232]. The effect of using conventional adjustment methods vs. causal modelling with respect to the size and strength of evidence for the association between current CD4 cell count/viral load and DM was not reported. However, presumably if associations had markedly altered in the sensitivity analysis with respect to these variables the authors would have reported this.

The Italian analysis undertaken by Spagnuolo et al. did not clearly indicate why values for nadir CD4 cell count and duration of known HIV infection were not reported in the multivariable results [229]. The methods section stated that, "*The multivariate model included characteristics with a p-value of ≤ .20 in the univariate analyses or factors…commonly associated with DM*". In the univariable analysis the HR for the association between nadir CD4 cell count and DM had a p value of 0.02 and the p value for duration of HIV infection was 0.003.

In the analysis undertaken by Rotger et al. the point estimate for current CD4 cell count was not provided. The p value was reported however (p=0.02). This p value would indicate that the 95% CIs for the IRR did not include one. However Figure 2 in the publication is a forest plot in which the IRR appeared to be almost exactly one with 95%CIs so small they were not visible [225].

8.8.8 Sensitivity analyses

The sensitivity analyses undertaken by studies are summarised in Table 8.19b. A number of the sensitivity analyses were not directly relevant to my associations of interest.

Table 8.19a: A comparison of characteristics of studies examining the association between HIV-related factors and DM

	My study	D:A:D (mostly EU) 2008 [158]	D:A:D (mostly EU) 2012 [226]	SHCS (Swiss) 2007 [224]	SHCS (Swiss) 2010 [225]	San Raffaele hospital (Italy) [229]	MACS (US) [124]	South Carolina Medicaid (US) [232]
STUDY CHARACTERISTICS								
Study Type	Collaboration	Clinical cohort		Clinical cohort		Clinical cohort	Interval cohort	Clinical cohort
Country/Region	Mostly EU	Mainly EU		Switzerland		Italy	USA	USA
Cohort type(s)	Clinical and Interval [a]	Clinical & Interval (11 cohorts)		Clinical (7 Sites)		Clinical (1 site)	Clinical (4 Sites)	Insurance database
Analysis design	Cohort	Cohort		Cohort		Cohort	Cohort	Cohort
FU period	1980-2013	1999-2006	1999-2010	2000-2006	1999-2009	1991-2014	1999-2003	1994-2012
PYFU [IQR]≠	5.4 [2.3-10.9]	*4.0†*	5.2 [3.0-8.1]	*4.3†*	9.7 (8.6-9.9)	9.8 [4.3-16.3]	2.3 [1.1-3.0]	5.8 [1.9-8.8]
Study Overlap	D:A:D	My study SHCS		D:A:D		None	None	None
Publication Year	-	2008	2012	2007	2010	2017	2005	2014
DEMOGRAPHY								
Age: FU start	32 (26-39)	38 (33-44)	46 (41-53)^	38 (34-44)	40 (35-48)	35 (30-42)	46 (42-51)	39 [31-46]
Sex: Male	86%	74%	73%	69%	80%	78%	100%	57%
Ethnicity: Black	6%	10%	7%‡	11%	0%[c]	7%‡	14%‡	71%
Annual LTFU (cohort)	4%	5-8%	5-8%	-	-	-	27% [b]	-
DATA SOURCES								
Outcome	MR	CRFs		EMR/CRFs		EMR	Self-report & biannual lab tests	Medicaid database
Other	(E)MR/ registry	(E)MR/Interval data Registry		EMR/Resistance database/CRF		EMR	Self-report & biannual clinic visits	Registry

ICD-International Classification of Diseases (E)MR-(Electronic) medical records ‡Non-white †IQR/95%/SE CI not reported ≠Median PYFU (unless in italics which denotes mean) ^ baseline was the 1st time-point after D:A:D enrolment when all DM risk factors were present. [a] PRIMO and SEROCO cohorts are interval cohorts [b] Total lost to follow-up during DM study follow-up [c] Sub-population was all white, due to inclusion criteria for genetic study

Table 8.19b: A comparison of characteristics of studies examining the association between HIV-related factors and DM

	My study	D:A:D (majority EU) 2008 [158]	D:A:D (majority EU) 2012 [226]	SHCS (majority EU) 2007 [224]	SHCS (majority EU) 2010 [225]	San Raffaele Hospital (Italy) [229]	MACS (US) [124]	South Carolina Medicaid (US) [232]
OUTCOMES								
Both types/ Type-2 DM only	Both	Both		Type-2 only		Type-2 only	Both	Both (SA undertaken)
Case definition	Record & date in (E)MR	Definitive: fasting plasma glucose ≥7.0mmol/L on 2 occasions Probable: DM documented in clinical notes with date of diagnosis		Definitive: ≥7.0 (fasting) & >11.1 (non-fasting) mmol/L		American Diabetes Association Criteria‡	≥7 mmol/L (fasting) glucose or self-report (diagnosis or treatment)	ICD-9
Validation/ Review	No/No	Yes/Yes		Yes/Yes		Yes/Yes	No/No	No/No
Crude rate /1000 PYFU (95%CI)	1.9 (1.6-2.3)	5.7 (5.3-6.1)	4.2‡	4.4 (3.7-5.3)	-		47 (32-71)	11‡
Exposures of interest	A-E	A,C	B	C	B	A-D	C	B, D
POTENTIAL CONFOUNDERS								
Not captured	5,7-9,11-13	8, 9^,10-13	4,7^,11,13	9,11	2,4,7,10-13	8,11	2,4,7,8,10-13	2,11-13
Time-updated	0,6,10	0,6,7	0,5,6,8-10	0,5,7,8	5,6,9	0,9,12	-	5,6,8,9,10
Time-fixed	1-4	1-5	1-3	1-4,6,10,12,13	0,1,3	1-3*,4-7,10	0,1,3,5,6,9	0,1,3,4,7
Missing reported	For all	No	Overall	3,10	No	2,5,7,10	5	No
Missing >10%	3	-	-	-	-	2,5,7,10	-	-
STATISTICS								
Missing data	Complete case	Complete case		Complete case	Complete case	Complete case	Complete case	Complete case
Statistical model	Poisson	Poisson		Poisson	Poisson	Cox (competing risks)	Poisson and Cox	Marginal structural model
Method of variable selection	*A priori* & stepwise (<0.1)	*A priori (*for non-ART variables)	Backwards stepwise (<0.05)	*A priori*	*A priori*	p≤0.2 in univariable analysis & ART	Not stated	Backwards-stepwise p<0.08
Assumptions tested	Region as proxy for ethnicity	-	-	Linearity of associations	-	Proportional hazards	Linearity, proportional hazards (cubic splines)	Proportion hazards
Sensitivity analysis	Carry forward, calendar period, linearity	Inclusion:lipids, triglycerides & fat loss	Missing data	Stratification, time-updating, lagging ART	ART (various) HCV (inclusion)	None reported	ART (different measures)	Standard Cox model, Type2 DM only

A-duration of infection, B-current CD4, C-Nadir CD4, D-Current Viral Load, E-Duration of immunosuppression Potential confounders: 0-age, 1-sex, 2-mode, 3-ethnicity, 4-calendar period, 5-BMI/central obesity, 6-ART, 7-Smoking, 8-Hypertesion, 9-Lipids, 10-HCV, 11-Parental DM, 12- non-ART medication, 13-Co-morbidity *captured but not included as almost all population white

^captured but not evaluated in the relevant multivariable analysis ‡95%CI not reported

8.8.9 A comparison of findings across studies

Table 8.20a-d summarise the results for the associations between my exposures of interest and DM across studies. Generally there is little evidence of an association between HIV-related factors and DM, although results were conflicting for current CD4 cell count, with some studies finding evidence for an association (including D:A:D).

Table 8.20a: A summary of findings: The association between duration of HIV infection and DM before and after adjustment

		Duration of HIV infection				Factors accounted for in multivariable analysis*														
Study (Location)	**N/n (HIV positive/ events)**	**Point Estimate**	**Unit increase in years**	**Point Estimate (95% CI) p-value**		**Age**	**Sex**	**Mode**	**Ethnicity**	**Calendar Period**	**BMI**	**Central obesity**	**cART use**	**Smoking**	**Hypertension**	**HCV**	**Lipids**	**Current CD4**	**Nadir CD4**	**Current Viral Load**
				Univariable	**Multivariable**															
My study	8 166	IRR	Per 10	**2.14 (1.56-2.92)**	1.02 (0.68-1.52)	**√**	√	√	√	**√**	x	x	**√**	x	x	**√**	x	**√**	**√**	**√**
(EU/ Canada)	95			**<0.0001**	0.9															
D:A:D (mostly EU) 2008 [158]	32 437	IRR	Per year	≠	0.98 (0.96-1.00)	**√**	√	√	√	√	√	x	√	√	x	x	x	x	x	x
	744				0.09[a]												^		^	
San Raffaele Hospital (Italy) [229]	6 195	HR	Per 5	**0.85 (0.76-0.95)**	≠[†]	√	√	√	√	√	√	x	√	√	x	√	**√**	**√**	√	**√**
	235			**0.003**																

*By restriction, matching, adjustment, randomisation or found not to be associated with the outcome and then dropped **√**-Time-updated variables in bold # At baseline (ART start) ^captured but not included in the relevant multivariable analysis [a] Duration of known infection at D:A:D enrolment ≠ Not reported [†]Not reported, but unclear why this has not been reported as methods state all variables p≤0.2 in univariable analysis were included in multivariable model

Table 8.20b: A summary of findings: The association between nadir CD4 cell count and DM before and after adjustment

				Nadir CD4 cell count		Factors accounted for in multivariable analysis*														
Study (Location)	**N/n (HIV positive/ events)**	**Point Estimate**	**Unit Increase Cells/µL**	**Point Estimate (95% CI) p-value**																
				Univariable	**Multivariable**	**Age**	**Sex**	**Mode of Infection**	**Ethnicity**	**Calendar Period**	**BMI**	**Central Obesity**	**cART use**	**Smoking**	**Hypertension**	**HCV**	**Lipids**	**Current CD4**	**Current viral Load**	**Duration infection**
My study	8 161	IRR	100	0.92 (0.83-1.02) 0.1	1.07 (0.96-1.21)	**√**	√	√	√	**√**	x	x	**√**	x	x	**√**	x	**√**	**√**	**√**
(EU/ Canada)	95				0.2															
MACS (US) [124]	229	IRR	≤300	≠	1.67 (1.00-2.80)	√	√ [¥]	x	x	x	√	x	√	x	x	x	x	x	x	x
	28		>300		≠															
SHCS 2007 (Switzerland) [224]	6 513	IRR	<200	1.56 (0.86-2.82)	0.96 (0.46-2.02)	√	√	√	√	x	√ [b]	√	x ^	√	√	x ^	x	√ [c]	x	x
	123		200-499	1.06 (0.57-1.97)	0.98 (0.49-1.93)															
			≥500	1	1															
D:A:D (mostly EU) 2008 [158]	32 437	IRR	50	≠	0.98 (0.96-1.00)	**√**	√	√	√	√	√	x	√	√	x	x	x ^	x	x	x ^
	744				0.06															
San Raffaele Hospital (Italy) [229]	6 195	HR	100	**0.89 (0.80-0.98) 0.02**	≠†	√	√	√	√	√	√	x	√	√	x	√	**√**	**√**	**√**	**√**
	235																			

*By restriction, matching, adjustment, randomisation or found not to be associated with the outcome and then dropped **√**-Time-updated variables in bold ^ captured but not included in the relevant multivariable analysis ≠Not reported †unclear why as methods state all variables p≤0.2 in univariable analysis were included in multivariable model ¥-All male cohort [a] Duration of known infection at D:A:D enrolment [b] omitted from the model due to collinearity with central obesity [c] baseline CD4 in the model where IRRs were reported

Table 8.20c: A summary of findings: The association between current CD4 cell count and DM before and after adjustment

Current CD4 cell count (time-updated)						Factors accounted for in multivariable analysis*														
Study (Location)	**N/n HIV positive/ events**	**Point Estimate**	**Unit Increase Cells/µL**	**Point Estimate (95% CI) p-value**		**Age**	**Sex**	**Mode**	**Ethnicity**	**Calendar Period**	**BMI**	**Central obesity**	**cART**	**Smoking**	**Hypertension**	**HCV**	**Lipids**	**Nadir CD4**	**Viral Load**	**HIV duration**
				Univariable	**Multivariable**															
My study (EU/ Canada)	8 161 95	IRR	100	1.00 (0.931.07) 0.9	1.01 (0.94-1.09) 0.7	**√**	√	√	√	**√**	x	x	**√**	x	x	**√**	x	**√**	**√**	**√**
SHCS 2010 [225]	644 94	IRR	√	**≠**	**≠** **0.02**	√	√	x	x	x	**√**	x	**√**	x	x	x	**√**	x	x	x
D:A:D 2012 (mostly EU) [226]	16 632 376	IRR	<200 ≥200<350 ≥350	**≠**	**1** **0.52 (0.36-0.77)** **0.51 (0.37-0.69)** **<0.001**	**√**	√	√	√	x	**√**	x	**√**	x	**x**	**√**	**√**	x	**√**	**√**
San Raffaele Hospital (Italy) [229]	6 195 235	HR	100	0.95 (0.89-1.10) 0.1	**0.91 (0.84-0.99)** **0.03**	√	√	√	√	√	√	x	√	√	x	√	**√**	**√**	**√**	**√**
South Carolina Medicaid (US) [232]	6 816 491	HR	<200 200-499 ≥500	**≠**	Eliminated >0.05	√	√	x	√	x	**√**	x	**√**	x	**√**	**√**	**√**	x	√	x

*By restriction, matching, adjustment, randomisation or found not to be associated with the outcome and then dropped. **√**-Time-updated variables in bold ≠ Not reported

[a]Duration of known infection at D:A:D enrolment

Table 8.20d: A summary of findings: The association between current HIV viral load and DM before and after adjustment

Current HIV viral load (time-updated)						Factors accounted for in multivariable analysis*														
Study (Location)	**N/n HIV positive/ events**	**Point Estimate**	**Unit Increase Copies/mL**	**Point Estimate (95% CI) p-value**		**Age**	**Sex**	**Mode of**	**Ethnicity**	**Calendar Period**	**BMI**	**Central Obesity**	**cART use**	**Smoking**	**Hypertension**	**HCV**	**Lipids**	**Current CD4**	**Nadir CD4**	**Duration of**
				Univariable	**Multivariable**															
My study	8 161	IRR	Log_{10}	0.86 (0.74-1.00)	1.09 (0.93-1.26)	√	√	√	√	√	x	x	√	x	x	√	x	√	√	√
(EU/ Canada)	95			0.05	0.3															
South Carolina Medicaid (US) [232]	6 816	HR	Log_{10}	≠	Eliminated	√	√	x	√	x	√	x	√	x	√	√	√	x	√	x
	491				>0.05															
San Raffaele Hospital (Italy) [229]	6 195	HR	Log_{10}	**1.26 (1.11-1.43)**	**≠**	√	√	√	√	√	√	x	√	√	x	√	√	√	√	√
	235			**<0.001**																
			≥50 vs<50	**1.52 (1.10-2.10)**	**2.00 (1.41-2.84)**															
				0.01	**0.0001**															

*By restriction, matching, adjustment, randomisation or found not to be associated with the outcome and then dropped **√**-Time-updated variables in bold [a] Duration of known infection at D:A:D enrolment ≠not reported

9 Discussion

9.1 Summary of Findings and Implications

9.1.1 Introduction

PLHIV experience higher rates of SNAEs than the general population, but the contribution of HIV (or related factors) to this higher rate is unclear [117, 167]. The aim of my study was to explore the association between HIV-related variables and individual SNAEs incidence before and after adjustment. I had sufficient data to examine fractures, MI and DM as outcomes. I investigated whether duration of HIV infection was associated with these events. I also explored whether levels of current HIV viraemia and/or immunosuppression (current and nadir CD4 & time spent with CD4 ≤50/100/200 cells/µL) were associated with each SNAE. My study was observational and therefore could not establish causality, but I adjusted for established risk factors where available.

9.1.2 Fractures: Summary of my study findings

There was little evidence that my exposures of interest except prior AIDS were associated with fractures. All relevant exposures (with the exception of current CD4 count) were associated with fracture incidence in the univariable analysis (Table 6.20). After adjustment for potential confounders however, evidence of an association remained only for prior AIDS (~50% increase in incidence rate). The final model included demographic factors (including current age), time-updated ART (TDF, PI & "other"), HCV, prior AIDS, cohort and calendar period. An interaction term was also between sex and age. The results were robust to sensitivity analysis.

9.1.3 Fractures: Implications of my study findings

My study did not find evidence to suggest fracture rates are independently associated with HIV-related factors (except prior AIDS). Increased fracture incidence in PLHIV compared to the general population may therefore largely reflect exposure to ART (PIs and TDF), HCV infection or other risk factors.

If fragility fractures are causally associated with HIV infection however, then a likely mechanism would be through HIV's effect on BMD. In-vitro studies have shown that the virus affects bone turnover [290]. It increases osteoclast activity leading to bone resorption and decreases new bone formation by osteoblasts [317, 318, 350]. As this process is progressive, a time-lag between exposure and increased fracture incidence seems plausible. Although my follow-up was not

insubstantial it may still have been insufficient to detect associations. I discuss potential sources of bias and their possible impact on my findings further in Sections 9.2 & 9.3.

9.1.4 Myocardial Infarction: Summary of my study findings

After adjustment (for current age, sex, ART, calendar period and cohort) duration of HIV infection and prior AIDS remained independently associated with MI incidence. For each additional 10 years of HIV infection, MI incidence increased by about 90%. Those with a prior AIDS diagnosis were almost 70% more likely to experience an MI. All findings were robust to sensitivity analysis.

In my univariable analysis I also found evidence of an association between both nadir CD4 cell count and duration of immune-suppression ≤200/≤100 cells/μL with risk of MI (Table 7.4). These associations were attenuated and no longer statistically significant after adjustment. Evidence of an association between MI incidence and current HIV viral load, current CD4 cell count and duration of immune-suppression ≤50 cells/μL was lacking in both univariable and multivariable analysis.

9.1.5 Myocardial Infarction: Implications of my study findings

These results suggest that those with longer duration of HIV infection might be at increased risk of MI. I found that this association was independent of current age and the severity and chronicity of immune-compromise or current levels of viraemia. These findings could be due to confounding or other bias (discussed in Sections 9.2 & 9.3). However, if this association is causal then it has implications for the effective assessment of future MI risk in HIV-positive individuals. It may be beneficial to initiate risk reduction strategies (e.g. statin prescription or life-style change programmes) at lower thresholds of other risk factors in people with long-standing HIV infection than in the general population or those more recently infected.

I found that this association was independent of HIV viral load or CD4 metrics. There is, in the case of MI, a plausible mechanism by which duration of HIV infection could impact MI incidence in those on effective treatment (i.e. with low viral load and high CD4). Coronary artery disease, of which MI is a sequela, is now thought of as a consequence of chronic inflammation, endothelial dysfunction and aberrant coagulation [351, 352]. Those with HIV are commonly chronically inflamed. This is due, at least in part, to the extensive and permanent depletion of CD4 cells from the gut lymphoid tissue during very early HIV infection [170, 171].This results in large numbers of microbes entering the circulation via the gut which leads to marked up-regulation of inflammatory pathways [171]. There is evidence that higher than normal levels of inflammation commonly persist despite effective ART, which could explain my findings [353, 354]. If HIV is causally associated with MI then it is likely to be

mediated through the development of progressive atherosclerosis over time as a result of chronic ineffective upregulation of inflammatory pathways and lipid dysregulation [325].

9.1.6 Diabetes Mellitus: Summary of my study findings

There was little evidence from my study that HIV-related factors are associated with DM. In contrast to the situation for fractures and MI, prior AIDS was not associated with DM. Duration of HIV infection and duration of immunosuppression ≤200 cells/μL were associated with DM incidence in the univariable analysis. However, both rate ratios were attenuated and evidence of an association between either variable and DM was lacking after adjusting for current age.

These findings were robust to all sensitivity analyses with one exception (Table 8.9). In univariable sensitivity analyses which censored follow-up for each patient at the first gap in CD4 cell count measurements of >1 year, all three measures of duration of immune-suppression (≤200/100/50 cells/μL) were found to be associated with elevated rates of DM. Although these associations were attenuated by adjustment for current age, evidence of an association between duration of immune-suppression and DM remained after adjustment. I explored these results further and found that the sensitivity analysis censored follow-up where long durations of immunosuppression were generated (by carrying last observation forward) but no events occurred.

9.1.7 Diabetes Mellitus: Implications of my study findings

I found very little evidence that HIV-related factors are associated with DM. With respect to the sensitivity analysis which conflicted with the main results, its findings suggest two possible interpretations. Firstly, that under-reporting of events occurred preferentially in those with low CD4 cell counts who failed to attend clinic. In this scenario the sensitivity analysis removed bias and revealed a genuine association. The second possibility is that the sensitivity analysis introduced bias by removing long periods of genuine severe immuno-suppression where no events occurred.

9.1.8 Limitations across outcomes

My results may reflect the true nature of the relationships between my exposures of interest and outcomes, but the influence of bias cannot be ruled out. In Section 9.2 and 9.3 I discuss the characteristics of my study in the context of bias and describe my attempts to minimise its effects of my results. Studies often did not report both univariable and multivariable results so it was often not possible to see if adjustment may have impacted findings.

9.2 Study strengths

9.2.1 Well-estimated HIV seroconversion date

A major strength of my study was that all individuals enrolled in CASACADE have a well-estimated time of HIV seroconversion. This permitted me to calculate duration of HIV infection, nadir CD4 cell count and duration of immune-suppression with reasonable accuracy. Most previous studies examining these variables [124, 174, 176, 178, 183, 186, 187, 199, 203, 224] had missing information from seroconversion to the first known positive HIV test [210]. An additional advantage of my sero-incident analysis was that individuals who died early in the course of HIV infection were more likely to have been included and so survivor bias was less likely than in previous studies (which were largely in sero-prevalent cohorts) [355].

9.2.2 Pre-planned data pooling

My study pooled data from multiple cohorts across countries with similar socioeconomic and demographic characteristics. This was advantageous because it: was cost effective, increased study power, permitted the analysis of rare outcomes and enhanced generalisability [356]. My study is likely to have been at a reduced risk of bias due to study heterogeneity and lack of data harmonisation compared to studies which relied on post-hoc data pooling (e.g. individual patient data or standard meta-analyses) [357].

9.2.3 Comparability of exposure groups

It is not uncommon in observational studies for exposed and unexposed individuals come from different populations. This can introduce bias because the two populations are likely to differ with respect to other characteristics in addition to exposure status [312]. All individuals in my study came from the same population regardless of exposure status (CASCADE). In fact, during my analyses individuals could contribute time to different exposure categories as variables were time-updated.

9.2.4 Population representativeness

Groups commonly excluded from clinical trials including the elderly, those with comorbidities and pregnant women were included in my study (see Chapter 5: Data Overview) [358]. My study population was therefore potentially more representative of PLHIV in high-income countries than individuals commonly included in HIV RCTs in a similar setting.

9.2.5 Temporality

Due to my study's longitudinal nature I was able to determine (from the point at which each cohort started capturing SNAE data) the timing of an exposure relative to the timing of the outcome (temporality). The risk of reverse causality was therefore minimised.

9.2.6 Repeated measures

All cohorts in my study collected repeated measures for CD4 cell count, HIV viral load and ART as they changed over time. I generated time-updated exposure variables in my analyses. This permitted me to explore whether current CD4 and HIV viral load, cumulative CD4, prior AIDS and nadir CD4 values were associated with individual SNAEs. I also undertook time-updated adjustment for potential confounders (age, ART and HCV). The cohort design of my study enabled me to calculate incidence rates and to explore multiple exposures and outcomes.
Many of the studies included in my literature reviews only included time-fixed values (commonly at enrolment or ART start).

9.2.7 Substantial medium to long term follow-up

For each of my analyses there was a substantial amount of medium to long-term follow-up. The percentage of follow-up accrued ≥10 years after HIV seroconversion was 25% for fractures 28% for MI and 32% for DM. All three analyses had a median follow-up of ≥ 5.3 [widest IQR, 2.3-10.9] years per person. It is plausible that there may be a marked time-lag between exposure and outcome in my study. Time-interval bias occurs when there is insufficient follow-up time for the effect of exposure on outcome to be observed [358].

9.2.8 Hard end-points

All three outcomes I examined were hard end-points. The use of surrogate end-points can lead to bias [359]. Composite end-points can increase study power but can lead to statistical challenges and introduce a potential source of bias [360]. A number of studies have examined SNAEs as a composite end-point [18, 167, 361, 362]. These studies assume the associations between exposures and the component outcomes are the same. Given the very different risk factors and mechanisms leading to the development of SNAEs this assumption is not very biologically plausible. When D:A:D explored viral-associated cancers, which are far more similar as a group than MI and liver disease for example, they found the size and direction of associations between CD4 metrics and each cancer differed by cancer type (Caroline Sabin personal communication).

9.2.9 Validity of exposure status

My study did not rely on self-report with respect to exposures of interest (duration of HIV infection, CD4 and HIV viral load). With the exception of mode of infection and ethnicity, variables I adjusted for in my multivariable models also did not rely on self-report. These comprised: age, sex, calendar-period, ART, HCV and prior AIDS. Recall bias can occur if individuals differ in the accuracy of their recall of true exposure status dependant on their outcome status [363].

9.3 Study limitations

9.3.1 Confounding: variables not captured

In general cohort studies provide a lower grade of evidence than RCTs, due to the likelihood of results being influenced by residual confounding [364]. However, some research questions such as my own cannot be answered using an RCT. Cohort study findings usually provide stronger evidence than case-control studies [358]. No other observational designs (ecological, cross-sectional or case-series) would have been suitable for answering my research question.

I was missing data on known predictors of my outcomes, which I discuss with respect to individual outcomes in Section 9.3.3. I had hoped to undertake a one-off data merger to capture additional potential confounder variables (smoking, hypertension, lipids and height/weight) but was unable to do so due to resource constraints at a cohort level. The information on these factors was captured by most cohorts. What information was being captured was highly variable however. Also data capture had, for most cohorts, been initiated long after cohort inception and was therefore incomplete. Smoking data relied on self-report, which can be biased [365]. A lack of data harmonisation would have prevented adequate adjustment due to these issues even if I had been able to capture relevant data [366].

The absence of adjustment for traditional risk factors not captured in CASCADE could have generate spurious positive findings or even spurious negative ones. Pack-years of smoking, in current smokers, has been shown to have a (somewhat concaved) dose-response with atherosclerotic cardiovascular disease [367]. Some or all of the dose-response association I observed between duration of HIV infection and MI could therefore be due to smoking. Prevalence of smoking in both the general population and PLHIV is declining [368, 369]. However, smoking is more common in PLHIV than the general population. The majority of individuals in my study lived in France. During the later-part of my study follow-up period 38 % of French PLHIV smoked regularly [370].

There are a number of other viruses, not routinely captured in CASCADE, including cytomegalovirus (CMV), which have been found to be associated with severe non-AIDS outcomes and death [142, 143].

9.3.2 Confounding: variables captured but data incomplete

Data on some variables were missing for some individuals. I adjusted for prior HCV- seropositivity (fractures and DM analyses) but these data were incomplete. I assumed that those with no HCV sero-positive test results were HCV sero-negative. My data were collected before direct acting antivirals (DAA) became widely available [371]. Residual confounding by HCV-seropositivity may have occurred. It appeared however, that cohorts with apparently incomplete data had only sent information on individuals with positive test results. In light of this I think that my assumptions regarding HCV-seropositivity were reasonable. HCV-seropositivity is associated with PWID, which is in turn associated with lower CD4/high HIV viral load and poorer outcomes [7]. If my adjustment was incomplete it could possibly have generate a spurious association between high HIV viral load/low CD4 and fractures/DM. As there was little evidence from my analysis that these were associated, the impact of possible residual confounding by HCV-seropositivity appears minimal.
Comprehensive data on ART exposure were available in CASCADE. However, records of ART treatment for some individuals appeared to be missing stop dates for some treatment changes. This may have led me to misclassify patient's current ART status. It is also likely I over-estimate cumulative ART exposure as a result. As most individuals who were not lost to follow-up continued ART once they have started it, the effect of this misclassification may have been minimal, but it is difficult to predict its possible effect.

9.3.3 Outcome specific confounders

It is not well-established which factors are associated with my exposures of interest, but much previous work has been done on establishing risk factors for my outcomes.

For my fractures analysis data on some components of the FRAX score/Q Fracture risk prediction equations were not captured in CASCADE. These comprised: weight, height, previous fracture/falls, parental hip fracture, smoking, steroid use, rheumatoid arthritis, secondary osteoporosis, alcohol abuse, asthma, CVD and malabsorption [314, 372]. These risk factors for fragility fracture could have acted as confounders in my analysis. Smoking has been associated with changes in CD4 cell counts which may vary by ethnicity [373]. Prednisolone and BMI have been associated with increased CD4 cell count and reduced rates of disease progression [374, 375]. Alcohol has been associated with

reduced CD4 cell count and accelerated disease progression [376]. If immunosuppression (current or cumulative) is causally associated with fractures, then some (or all) of its effect could be mediated through BMI, so it should not probably be adjusted for.

Framingham risk factors, which are relevant to my MI analysis, but not captured in CASCADE comprise: BMI, previous smoking, cholesterol (total and HDL) and systolic blood pressure. Some known risk factors are excluded from the Framingham risk equation, these are: DM, chronic kidney disease, symptomatic carotid artery disease (CAD), clinical coronary heart disease (CHD), abdominal aortic aneurysm (AAA) and peripheral arterial disease (PAD). These factors are known as "risk equivalents" because they are considered to carry the same risk as having had a prior MI (i.e. a 10 year risk of MI or coronary death of >20%) [377, 378]. Other factors not included in Framingham but known to affect CVD (and therefore MI) risk and are included in QRisk3 [213]. I was missing the following QRisk3 variables: family history of angina/MI <60 years of age, atrial fibrillation, taking blood pressure treatment, prior migraines, systemic lupus, severe mental illness, use of antipsychotics or steroids and erectile dysfunction. All these factors increase the risk of primary MI and could act as confounders in any analysis if they are associated with exposures. I was unable to adjust for DM in my MI analysis due to data on each not being from the same cohorts.

Similarly, in the DM analysis I had no data on QDiabetes (a well-validated DM risk prediction equation) risk factors [227]. These comprised: smoking, deprivation, family history, prior CVD, blood pressure, learning disability, regular steroid use, severe mental illness, gestational diabetes, polycystic ovarian disease, statin use, atypical antipsychotic use and BMI.

Any study would have to be extremely well-powered to permit adjustment for even a small portion of these additional factors.

9.3.4 Loss to follow-up: risk of informative censoring

Cohort studies are prone to selection bias due to loss-to-follow-up [379-381]. Some authors have suggested a percentage threshold for loss at which major bias in the findings might be anticipated [382]. However, exploratory analysis suggests that it is whether the loss is informative [379, 380]. If those lost are lost (become missing from follow-up) not at random (MNAR) with respect to exposure and outcome status then even low levels of loss can adversely impact the internal validity of the results [379]. Conversely high levels of loss (up to 60%) can still produce valid results if loss-to follow-up occurs at random (MAR) or completely at random (MCAR)[379]. My fracture and MI analyses included data on all seroconverters in the FHDH (the largest contributor to my study). The study has

been capturing data on SNAEs since cohort inception in 1989. Follow-up in my study ended in 2013. A large number of individuals ever enrolled and potentially eligible for inclusion in my analysis have been lost to follow-up in the intervening years (50%). If those lost to follow-up in my analysis differed with respect to the relationship between exposure and outcome compared to those who remained in follow-up, this could bias results. Informative censoring commonly biases results towards the null [383]. A number of methods have been developed to attempt to address informative censoring including the use of inverse probability of censoring weights (IPCW) [380, 384]. This was considered beyond the scope of my thesis. I would have had to firstly split my data into equal periods of follow-up (typically of 1, 2 or 3 months) which would have been non-trivial as I had set up the data using periods that corresponded to changes in time-dependent covariates. I would then have had to investigate predictors of loss to follow-up over time by cohort (probably needing further information from cohorts on classification and reasons for loss to follow-up), to then derive separate censoring models for each cohort over time. Finally I would have needed to rebuild and refit my models using weighted Poisson models [385].

9.3.5 Time-interval bias

I discuss time-interval bias in section 9.2.6 [358]. Despite the substantial follow-up time in my study it is possible it was still insufficient. If there is a very long time-lag between my exposures of interest and outcomes (i.e. several decades), then more long-term follow-up may be needed to detect associations.

9.3.6 Lower than anticipated study power

I undertook a survey of CASCADE cohorts to determine what data cohorts were capturing on SNAEs and approximate numbers of events. It appeared from the information that I was provided with, that there were far more events than I ended up with in my analyses. This was because a number of cohorts provided me with the numbers of SNAEs of each type they had in their entire cohort as opposed to the number of events in individuals enrolled in CASCADE. Most collaborating cohorts are sero-prevalent and only have a small percentage of individuals with a well-estimated time of HIV seroconversion (a requirement for CASCADE). This meant that my study was not as well powered as I had initially anticipated. I therefore may have failed to detected small differences in outcome incidence between exposure groups that genuinely were present

9.3.7 Survivor bias

My study analysed data from a sero-incident cohort collaboration, which theoretically reduced the risk of survivor-bias as previously discussed (Section 9.2.1). However, unlike FHDH (my largest contributor) some of the collaborating cohorts did not start capturing relevant outcome data until a number of years after cohort inception. I started follow-up for individuals from these cohorts on the date event capture commenced. This prevented individuals accruing follow-up when they could not have had an event recorded. However, if patients had died before this point then they would not have been included in my analysis. This could have led to some survivor-bias operating in my analysis.

9.3.8 Inclusion criteria: A lack of a HIV-negative comparison group

Unlike some HIV cohorts [293, 386, 387], CASCADE does not include an HIV-negative comparison group. I could not therefore compare event rates between PLHIV and a group of otherwise similar individuals without HIV.

9.3.9 Inclusion criteria: A well-estimated time of HIV seroconversion

Individuals who actively seek healthcare have been found to have better outcomes in terms of both morbidity and mortality [388, 389]. Those engaging in regular HIV testing are diagnosed earlier in the course of HIV infection compared to those who avoid testing. They are also likely to have higher CD4 cell counts and initiate ART earlier (especially in later calendar years due to changes in treatment recommendations). A well-estimated date of HIV seroconversion may have preferentially selected a sub-group of HIV-positive individuals with a better prognosis. This could have led to lower event rates in my study compared to the HIV-positive population as a whole.

9.3.10 Population: Access to free universal health-care

Individuals in my study had access to free universal healthcare. They therefore may be more representative of the general HIV-positive population across resource-rich countries than the subgroup of those enrolled in US private healthcare plans (which only include those who can afford them) [390]. Many observational studies examining these associations have been undertaken in the USA using healthcare plan databases [203, 207, 220]. It should be noted, that HIV cohorts have been found to under-represent marginalised groups such as recent migrants [391, 392].

9.3.11 Exposures of interest

My study did not capture CD8 T-cell counts, which would have been interesting to examine. Previous studies have found an association with low CD4/CD8 ratios and both non-AIDS events and non-AIDS death [362, 393].

As my data were largely from clinical cohorts, the frequency of CD4 cell count and HIV viral load measurements was dependent on the clinician's recommendation and therefore varied. Due to this variability it was not possible to distinguish missing values from those never taken due to non-attendance or clinical decision making (unless they were missing for very long periods).

During time-updated analysis, "current" values of CD4 and HIV viral load varied with respect to how recent they were. I used a last-value-carried-forward technique, but during sensitivity analysis I examined the effect of using this method.

9.3.12 Channelling-bias and time-dependent confounding

I only included current and cumulative measures of ART exposure in my analyses to adjust for their potentially confounding effects. My study was not designed to examine associations between ART and SNAEs. The associations I found between ART and my outcomes could be due channelling bias (also known as confounding by indication) [394-396]. Channelling bias may occur when medication is chosen by the clinician based on patient characteristics related to that patient's frailty, disease severity or co-existing illness. If these variables are not captured and adjusted for in the analysis, then channelling bias can occur. As these characteristics are not captured within CASCADE no direct adjustment can be undertaken to mitigate their influence.

Time-dependent confounding by prior exposure could also be influencing the apparent association between ART and outcomes in my study [397]. Time-varying confounding by prior exposure occurs when:

1. The past value of a confounder (A) predicts the current value of an exposure (B)
2. The current value of that confounder (A) predicts the outcome (C) and
3. The past value of the exposure (B) predicts the current confounder value (A)

Such a scenario has been identified by other researchers with respect to adjustment for CD4 cell count when examining the effect of exposure to ART (B) on outcomes [397].

9.3.13 Collider bias

Observational researchers commonly attempt to reduce bias at the analysis stage by adjusting for potential confounders if variables meet certain criteria [398]. These standard methods are well-established, but have limitations. Spurious associations between exposure and outcome can be generated where none genuinely exists. This occurs if adjustment is undertaken for a variable which is causally influenced by both the exposure and the outcome [398, 399]. The phenomenon is called collider-bias. To identify colliders directed acyclic graphs can be constructed to explore the relationship between variables [398]. The risk of collider bias in my study was likely to be low due to the lack of data available on potential confounders and the lack of evidence for associations between exposures and outcomes.

9.3.14 Case definitions

Insufficiently robust case definitions can lead to outcome misclassification [400]. This may be non-differential, i.e. independent of exposure status. It may however be differential, i.e. the likelihood of misclassification varies by exposure status. Traditionally the inclusion of false positives was thought to reduce findings towards the null. It is now believed it can bias findings in either direction [400, 401].

Misclassification is more common for outcomes which are difficult to diagnose such as acute MI. For example attempting to identify acute MI using clinical signs without additional tests has very low sensitivity and specificty [402]. Given that I was only able to generate a very basic case definition (document record of MI and date) without validation/review, my study is likely to have had some false positives and negatives. I undertook a sensitivity analysis, however, restricting analysis to individuals from cohorts with well-validated cases. The results were similar with respect to the size and direction of associations to that of the main analysis.

There are also two types of acute MI. Type-1 (T1MI) occurs as a result of atherosclerosis secondary to chronic infection and inflammation [403, 404].T2MI is not associated with atherosclerosis, but occurs when the heart's requirement for oxygen exceeds the amount supplied to it [405, 406]. In the general population T2MI commonly occur secondary to: operations, sepsis, arrhythmia, heart failure or anaemia [339, 407]. A US study found that about half of all MIs classified as probable/definite events by adjudication in those with HIV were acutally T2MI [285, 339]. These T2MIs mainly occurred as a result of sepsis or cocaine use [285]. T2MI occur most commonly in those who inject drugs and my study had a much smaller percentage of PWID (6%) than the North American

collaborative study (12%) [121]. The inclusion of T2MI is therefore likely to be less of an issue in my study that the US studies. Ideally I would have liked to be able to distinguish between the types and have examined them both together and seperately to see if observed assocaitions differed by type.

Again, ideally I would have liked to have examined fractures both overall and by type (low and high-impact). I was unable to distinguish between them. They have different aetiologies [408]. Fragility (low impact) fractures occur as a result of low bone mineral density. High-impact fractures commonly occur as a result of trauma (accidents and violence) and are associated with risk taking behaviours [408]. However, reductions in BMD (secondary to HIV or by other means) is likely to increase the risk of high impact fractures too. Therefore HIV might increase the incidence of both high impact fractures (due to both its association with risk taking and by its possible effect on BMD) and low impact fractures (through possible effect on BMD).

Type-1 (primarily an autoimmune disease) and type-2 (insulin resistance) DM have different causes and risk factors [227, 409]. I was primarily interested in type-2 DM, but was unable to distinguish between them. It is likely that there was only a small number of type-1 DM included in my DM outcome. T1DM tends to have a childhood onset and so would have been documented before the start of follow-up and therefore would have been excluded from my anlaysis.

9.3.15 Under-reporting

There is some evidence to suggest under-reporting of outcomes may have occurred in my study.

My outcome rates were lower than in some other studies [30, 121, 174], although these were predominantly US studies which are known to have higher rates of CVD and DM than Europe [410].

My rates varied markedly between cohorts and increased with calendar time. This could reflect differences in the distribution of risk factors and/or under-reporting by some cohorts. It also suggests either: improved reporting over time, a genuine increase in incidence in later calendar periods or greater test sensitivity latterly. Cohorts were heterogeneous in their methods of data capture (Section 3.1.8.1). Some cohorts used sophisticated national electronic medical records linked both to other clinics/hospitals and to registries (e.g. AHIVCOS). Some cohorts relied on annual follow-up forms. Individuals completing these forms were often required to tick a box if a relevant event had occurred during the previous year (e.g. UK Register). In some case that person had to manually check individual clinic records, which would be very time-consuming.

During the teleconferences I undertook, a number of PIs from participating cohorts (including AHIVCOS which had high-quality data) felt under-reporting was likely to be occurring in their cohorts

(see Section 3.1.8 for further discussion of this). Under-reporting could have led to bias, depending on whether or not the probability of an event being reported was associated with exposure status.

9.3.16 Changes in outcome classification over time

My study might have been affected by changes in the classification of outcomes over time or differences in classification between cohorts. MIs are challenging to diagnose. During my study's follow-up period the measurement of troponin became widely used, which is highly sensitive. This has helped improved the accuracy of diagnosis [411]. HbA1c measurement also became routine in the diagnosis and management of DM during study follow-up [412]. Sensitivity analyses where I restricted the follow-up to later calendar years (2005+) however found similar patterns of associations. There was no evidence that the size or direction of the results differed from the main analyses for any of my outcomes.

9.3.17 Competing risks

I have not yet adjusted for the competing risk of death due to other causes in my analysis. I plan to do this as part of future work. However, due to under-reporting of death (only 5% of individuals were recorded as having died during follow-up time in my analysis), so any correction of bias may be incomplete. The competing risk of death can lead to an under-estimation of the association between exposure and outcome [258].

9.4 Generalisability

Individuals included in my study were all from high-income countries. All cohorts were based in Western Europe with the exception of one cohort from Canada. The larger regional and national cohorts contributing data to CASCADE (such as FHDH) are thought to be representative of HIV-positive patients receiving care in those countries [391].

Only a small percentage of those enrolled in most contributing cohorts (with the exception of SEROCO, PRIMO and the UK Register) are eligible for inclusion in CASCADE. My study population was therefore a subset of individuals from these cohorts for whom a well-estimated HIV seroconversion date could be established. This preferentially selected patients who were older, male (predominantly MSM) and white (as outlined in Section 9.4.1). White MSM (unlike heterosexual women or PWID of either sex) have been found to have a life-expectancy similar to that of the general population in recent years [11].

So my findings are only directly applicable to high-income countries and largely apply to white MSM. However, the size and direction of the associations I found between my exposures of interest and

outcomes were similar when I stratified by sex or ethnicity. It is therefore possible that my findings are applicable to other groups. However, the percentage of women, non-white individuals and those who acquired infection through MSW or PWID routes are relatively small. Therefore my study may lack the power to detect differences in the nature of associations between exposures and outcomes in these groups relative to white MSM.

9.5 My findings in relation to those of other studies

9.5.1 Fractures

There is some evidence that HIV affects bone turnover [290] by increasing resorption and decreasing new bone formation [317, 318, 350]. This would lead to progressive loss of BMD, a key risk factor for (primarily fragility but possibly also high-impact) fractures [413]. Previous studies have shown PLHIV experience more fractures of both types (low and high impact) than the general population [119]. Fracture incidence has been found to be bimodal in its distribution with peaks in young men and older women [414].

Other possible reasons for increased fracture risk in PLHIV include exposure to ART, HCV and established fracture risk factors. The SMART and START RCTs, did not look at fractures, but found that the trial arms which took more ART (continuously in the case of SMART or initiating at higher CD4 counts in the case of START) experienced higher rates of BMD loss when compared to those randomised to the arms which received less ART [415, 416]. A number of ART drugs have been associated with fractures. TDF and PI have been associated with lower BMD and increased fracture incidence [154, 417-419].

The evidence from other studies that my exposures of interest are associated with fracture incidence is conflicting and not compelling. Only two studies examined HIV infection duration as an exposure, but neither my study nor the Swiss analysis found evidence for an association after adjustment [185]. In my study adjusting for current age was responsible for pushing findings towards the null.

A number of studies examined nadir CD4 cell count. The majority of point estimates (including mine) were close to one after adjustment or the variable was dropped during multivariable model selection due to lack of evidence for an association [186, 187, 196]. The HOPS study and the Boston study (95%CIs wide) did find evidence of association, but neither adequately adjusted for ART [182, 197]. This hypothesis is supported by my study; adjustment for time-updated ART (including TDF, PI

and other ART) attenuated my IRR and there was no longer much evidence of an association between nadir CD4 cell count and fractures.

Studies which included current CD4 cell count in their analyses also produced conflicting results. The Swiss study was of high-quality and did find a small inverse association. VACS-VC (also high quality), my study and WIHS found no association and all had point estimates which were very close to one with narrow 95%CIs [132, 184]. The Australian case-control study found a large association, but unlike the other studies used a binary cut off (<200 vs ≥200 cells/μL) [178, 185].

No studies found that higher viral load was associated with increased fracture incidence. VACS reported an inverse association after adjustment however [132, 184]. I also found a similar sized association to the VACS in my univariable analysis, but this was attenuated and no longer significant after adjustment for ART. The VACS results may result from residual confounding by ART. Results for studies which examined prior AIDS were also conflicting, with widely varying point estimates.

Results across studies were commonly not directly comparable as different cut-offs were used and study methods were heterogeneous, which makes it difficult to draw conclusions. The likelihood of a causal association between my exposures of interest and fractures is evaluated in 9.5.4.

9.5.2 Myocardial Infarction

It is well established that PLHIV are at a ~60% increased risk of MI compared to the general population [118]. It is unclear what the causal contribution of HIV, ART, traditional risk factors and HCV-seropositivity is to the increased incidence. MI results from atherosclerosis due to inflammation, endothelial activation and faulty coagulation and HIV leads to up-regulation of these processes [351, 352]. Previous work has shown that HIV is associated with higher levels of C-reactive protein, D-dimer and interleukin-6 which are well-established markers of inflammation and clotting [420]. Higher levels of these biomarkers have also been linked to an increased risk of ischaemic heart disease [421]. HIV replication leads to marked changes in lipid metabolism which may further fuels atherosclerosis [325].

In addition to the possible effect of HIV infection, PLHIV appear to have an additional burden of CVD risk due to ART exposure [152, 155, 202, 422, 423]. The results of observational studies which have explored the associations between specific ART and MI/CVD require caution in their interpretation due to their risk of bias (time-varying confounding and confounding by indication) [424, 425]. However, a recent study attempting to address these issues using marginal structural models found an association between abacavir exposure and MI in line with previous studies which did not use

causal modelling [217]. Traditional CVD risk factors have been found to be predictive of risk in those with HIV [216, 219] including BMI, lipids, blood pressure and smoking.

Many of the studies examining MI were of high quality and large (D:A:D, NA-ACCORD, VACS-VC, FHDH) with end-point review and validation and time-updated adjustment for relevant potential confounders [121, 199, 202, 207, 211]. However, even when comparing results across these large high-quality studies there was little agreement with respect to the association between HIV-related parameters and MI.

Only two studies examined duration of HIV infection. I found an independent linear dose-response association with MI incidence, with rates doubling per additional decade of infection. In contrast the KP study found no evidence of an association [220]. A dose-dependent response pattern suggests a causal relationship [426]. It is possible however, that this association was due to the impact of one or more unmeasured confounders. Possible candidates include cumulative measures of: smoking [427], dyslipidaemia, microbial translocation [428], hypertension [429], uncontrolled inflammation [430] or low CD4/CD8 ratio [431]. However it seems likely that with the exception of smoking these variables may be on the causal pathway between HIV infection duration and MI. The KP study adjusted for time-updated lipids and hypertension, examined duration of infection as a categorical variable and calculated HIV duration time from the time of the first known positive test. If dyslipidaemia and hypertension are on the causal pathway then this could have pushed the KP study findings towards the null. The study did not report univariable findings so I was unable to see whether an association was found prior to adjustment. The KP study also used broad categories of duration of HIV infection which may not have captured the true nature of the association.

All studies except D:A:D examining nadir CD4 cell count as a linear variable found a small decrease in risk of MI in those with higher nadir CD4 cell counts after adjustment (around 10% per 100 cell/µL increase) [199, 202, 203, 220]. However, results from neither my study nor the Boston Hospital study reached statistical significance [203]. Despite study heterogeneity findings were similar across publications. My study, FHDH and KP found a dose-response effect, but D:A:D did not. As the reported effect size across studies was small, it appears that nadir CD4 cell count is unlikely to be of major clinical significance in the development of MI.

In the KP study, due to incomplete ART data, the main analysis of PLHIV adjusted only for prior ART exposure at baseline (any/none) [220]. Individuals with lower recorded nadir CD4 cell counts might be expected to be more likely to have started ART. During the era of the study (1996-2009) many asymptomatic individuals did not start treatment until their CD4 cell count had fallen to <200

cells/µL (or ≤350 cells/µL in later calendar years) [432]. So the observed association may be due to confounding.

Reported associations with current CD4 cell counts were conflicting, and interpretation was difficult as most studies examined the variable categorically, but used various cut-offs. My study and the KP study did not find an association, but this was not due to a lack of power, because confidence intervals were narrow [220]. Where categorical analysis was undertaken, those with current CD4 cell counts ≤200 cells/µL were at increased risk (with the exception of D:A:D) [121, 207, 211, 221]. The same lack of agreement across findings was true for current HIV viral load. Only my study and D:A:D examined duration of immune-suppression and neither found an association [211]. I evaluate the likelihood of a causal association between HIV related factors and MI in Section 9.5.4.

9.5.3 Diabetes Mellitus

There is somewhat conflicting evidence regarding the relative prevalence and incidence of DM in PLHIV compared to that of the general population [123, 134, 232]. Exposure to ART has been associated with increased DM incidence, especially PI exposure [153, 223, 433]. PIs directly affect glucose metabolism [434] whilst certain NRTIs (notably stavudine) are thought to indirectly influence it [158]. Demographic factors (age and ethnicity) and traditional risk factors (including hypertension, obesity-especially abdominal fat, dyslipidaemia, smoking, hypertension and statin use) also heighten the risk of glucose intolerance and diabetes in those with HIV [226, 227, 435].

The studies which examined the association between both duration of HIV infection (three studies) and nadir CD4 count (five studies including D:A:D which had 744 events) and DM incidence did not find evidence for an independent association. Point estimates were very close to one and 95%CIs relatively narrow [158, 229]. There was some evidence of an association between both current CD4 cell count and current HIV viral load and DM including from D:A:D (for CD4). The size of the effect across studies was variable however, despite narrow 95%Cis, and it was therefore difficult to draw meaningful conclusion regarding these associations [226, 229].

9.5.4 Assessment of a likely causal association between my exposures and outcomes of interest

A number of methods have been developed which can be applied to observational studies to evaluate the quality of their evidence. More than fifty years ago Austin Bradford-Hill published a set of criteria to aid in determining the likelihood of causality in observational studies, which are still in use today [426]. More recently the Grading of Recommendations Assessment, Development and

Evaluation (GRADE) method has been developed [364]. It primarily focuses on systematically grading the strength of evidence for systematic reviews, clinical guidelines and public health interventions. Its components are useful to consider with respect to my research question however [364].

Table 9.1 summarises key components in the assessment of causality applied to my fractures analysis.

Table 9.1: An assessment of evidence for causality between HIV-related factors and fracture incidence

Factors	Characteristics	Strength of evidence of causal relationship
Study type	All cohort-studies (one case-control)	Starting point: Low grade evidence
Point estimate strength/consistency	-Highly variable across studies for all exposures -High quality analyses tended to show small effect size	These characteristics reduce the likelihood of a causal relationship
Imprecision	Most studies had narrow 95%CIs	Reasonable precision
Dose response	Not evident in most studies	
Adequate adjustment for likely confounders	Most studies either had: -Missing data on confounders -Adjustment was only partial	Results could be influenced by residual confounding
Directness	-Applicable population -Relevant exposure -Relevant hard end-point	Minimal issues with indirectness
Risk of publication bias	-Two smaller studies had strongest positive results -Point estimates frequently not reported if p≥0.05	Possible publication bias
Overall		**Low grade evidence**

GRADE rates cohort-studies as low grade evidence by default [436]. The quality of evidence can be upgraded in four ways three of which are applicable to my study: if the effect size is large, there is a dose response or all plausible confounding would push findings towards the null. The strength of evidence can be down-graded if there is: a high risk of bias, inconsistency, indirectness, imprecision or a high risk of publication bias [437].

I was not able to generate summary measures of effect due to study heterogeneity for all three outcomes. However, the size of the observed effect was small across all exposures of interest for fractures with the exception of two studies. Both these studies had ambiguous methods and one of

these studies (Alfred Hospital) was of a case–control design which is considered to be lower quality evidence by GRADE [178, 197] .

Table 9.2 summarises my assessment of the likelihood of causal associations between my exposures of interest MI incidence. There was more evidence that HIV-related variables might be causally associated with MI incidence than there was for fractures, but the strength of evidence was still low. For nadir CD4 the size of the point estimates were broadly similar across studies and there appeared to be a dose-response, so for nadir CD4 there was moderate evidence of an association.

Table 9.2: An assessment of evidence for causality between HIV-related factors and MI incidence

Factors	Characteristics	Strength of evidence of causal relationship
Study type	All cohort-studies (except FHDH nested case-control)	Starting point: Low grade evidence
Point estimate strength/consistency	-Highly variable across studies (& not due to lack of study power) for all exposures except for nadir CD4	These characteristics reduce the likelihood of a causal relationship except for nadir CD4
Imprecision	Most studies had narrow 95%CIs	Reasonable precision
Dose response	-Evident for nadir CD4 and infection duration (my study)	-Increases strength of evidence for nadir CD4 -Duration of HIV infection is only from one study
Adequate adjustment for likely confounders	Most studies: -Missing data on confounders or -Adjustment only partial	-Results could be due to residual confounding
Directness	-Applicable population -Relevant exposure -Relevant hard end-point	Minimal issues with indirectness
Risk of publication bias	-Point estimates frequently not reported if p≥0.05	Possible publication bias
Overall		**Moderate evidence for nadir CD4 cell count & low grade for other variables**

Table 9.3 summarises the strength of evidence for a causal association between my exposures of interest and DM incidence. The overall strength of evidence across studies for a casual association between my exposures of interest and DM was very low.

Table 9.3: An assessment of evidence for causality between HIV-related factors and DM incidence

Factors	Characteristics	Strength of evidence of causal relationship
Study type	All cohort-studies	Starting point: Low grade evidence
Point estimate strength/consistency	-Consistent lack of association for duration of HIV infection and nadir CD4 cell count -Highly variable across studies for current CD4 and viral load	These characteristics reduce the likelihood of a causal relationship
Imprecision	Most studies had narrow 95%CIs	Reasonable precision
Dose response	-Not evident in most studies across exposures	
Adequate adjustment for likely confounders	Most studies: -Missing data on confounders or -Adjustment only partial	Results could be due to residual confounding
Directness	-Applicable population -Relevant exposure -Relevant hard end-point	Minimal issues with indirectness
Risk of publication bias	-Point estimates frequently not reported if p≥0.05	Possible publication bias
Overall		**Very low grade evidence**

9.5.5 Lessons Learnt, Recommendations and Possible Further Work

Table 9.4 summarises key findings, limitations and possibilities for further work.

9.5.5.1 *Lessons Learnt*

I initially considered and requested information on a large number of non-AIDS events from cohorts. In retrospect, I ideally would have made preliminary estimates of the likely incidence of events based on knowledge of cohort follow-up and population risks. I could have then focused on a smaller number of outcomes from the beginning.

I would have requested fracture location where available. This would have permitted me to distinguish between likely fragility fractures and traumatic ones. I could have then subsequently undertaken a sensitivity analysis to see if results were similar by fracture type. I would also have requested what type of DM had been diagnosed where available, although adult onset type-1 is rare.

In hindsight I would have considered the feasibility of conducting a nested case-control study within CASCADE for MI. The design has a number of attractive features for answering my research question in CASCADE for this outcome. This approach would have allowed me to capture a greater level of information on a smaller number of cases and controls. I could possibly have used more robust case definitions and well-validated events, reducing misclassification [285]. A nested case-control design also has the potential to reduce under-reporting bias. In addition, this design makes data capture of additional potential confounders less burdensome on cohorts as it only requires data for a small number of individuals. It is likely however, that cohort level resource constraints would still have precluded this approach.

My exposures of interest were rarely of interest in other studies. They were often only included in analyses due to their potential to confound the associations of interest. This led to relevant publications being hard to find. By using systematic search methods I paradoxically missed relevant publications due to a lack of correctly ascribed MESH terms in PubMed. I found subsequently that it was often more effective to select likely journals and search their sites directly. If I had known this at the beginning of my study, it would have saved me time.

9.5.5.2 *Recommendations for other researchers in this area*

Robust case definitions, especially for cardiovascular diseases including MI, are important as misclassification can commonly occur and can influence findings [211]. This is very important for MI where ICD codes or a mention of the event in medical records is known to be commonly unreliable

[285]. Cohorts and collaborations have become more aware of this in recent years. D:A:D, NA-ACCORD and COHERE collaborations have done much work in ensuring cases are better validated and implementing end-point review where feasible [121, 211, 285, 391]

Table 9.4: A summary of: work I undertook to meet my thesis aims, key findings, limitations and possible further work

	Key Findings	Major Limitations	Further Work
Systematic review of relevant literature	***Chapter 2*** -Lack of consistency in findings across studies for all exposure-outcome combinations	-Paucity of relevant studies -Exposures of interest rarely of 1^0 interest in published studies, so relevant publications hard to find -Heterogeneity of methods and adjustment across studies	Contact authors requesting: -Point estimates where not provided due to non-significance -Clarification of method ambiguities
Survey of collaborating cohorts: assessment of SNAE data capture quantity and quality	***Chapter 3*** **-**Sufficient data to examine fractures, DM and MI	-Questionnaire over-estimated event numbers -Lack of capture of relevant data preclude use of robust case definitions	
Analysis			
Fractures	***Chapter 6, 7 and 8*** -Lack of evidence of an independent association between my exposures of interest (except AIDS)	-Lack of adjustment for well-established risk factors which were not captured -Lack of robust case-definitions -Possible under-reporting (rates lower than most other studies) -Possible informative censoring -Adjustment for ART also was likely to adjust for its effect on my outcomes mediated through its effect on my exposures	-Explore methods to address potential informative censoring -Examine the association between both viraemia copy-years and CD4 slope and my outcomes -Explore the use of multiple imputation for missing covariate data -Undertake competing risk model to account for the competing risk of death
Myocardial Infarctions	-HIV infection duration ~doubles the rate of MI per additional 10 years -Prior AIDS independently increases MI incidence -Nadir CD4 was inversely associated with MI. The effect size was similar to other studies		
Diabetes Mellitus	-Lack of evidence of an association between relevant exposures and DM		

9.5.6 Conclusion

I believe this is the first study to examine the associations between HIV-related factors and fractures, MI and DM in a large sample of individuals with well-estimated HIV seroconversion dates. Using seroconverter data enabled me to accurately estimate duration of HIV infection and immunosuppression as well as nadir CD4 cell count.

I found that HIV infection duration was independently associated with a near doubling of MI incidence for each additional 10 years of infection. This association was independent of CD4 cell count, HIV viral load or current age. I was unable to adjust for smoking as this information was not captured in CASCADE. Smoking could have acted as a confounder in my analysis and thereby generated a spurious positive association between duration of HIV infection and MI. If the association I observed is causal however then individuals with long-standing HIV infection are at increased risk of MI, even if effectively treated. People who have been HIV positive for a number of years may benefit from risk-reduction (e.g. statins) at lower levels of established risk factors than those currently applied. The HIV-positive population is ageing and average duration of HIV infection is increasing. Atherosclerotic cardiovascular disease is becoming an increasingly important cause of morbidity and mortality in PLHIV. Therefore if we can find effective interventions there is a big potential to improve the future health of PLHIV.

I also found that having experienced a previous AIDS event increased MI and fracture risk (but not DM). Although statistical evidence for an independent association between nadir CD4 and MI was lacking in my study, the size of my association was similar to other studies where statistical evidence was apparent.

Appendices

Appendix A: AIDS-Defining Conditions

AIDS-Defining Conditions (in those ≥13 years of age) [3]

Candidiasis of bronchi, trachea, oesophagus or lungs

Cervical cancer, invasive

Coccidiodomycosis, disseminated or extra-pulmonary

Cryptosporidiosis, extra-pulmonary or chronic intestinal (for >1 month)

Cytomegalovirus retinitis (with vision loss)

HIV-related encephalopathy

Herpes simplex: chronic ulcers (for >1 month) or bronchitis, pneumonitis, or oesophagitis (onset at age >1 month)

Histoplasmosis, disseminated or extra-pulmonary

Isosporiasis, chronic intestinal (for >1 month)

Kaposi sarcoma

Lymphoma, Burkitt (or equivalent term)

Lymphoma, immunoblastic (or equivalent term)

Lymphoma, primary, of brain

Mycobacterium avium complex or Mycobacterium kansasii, disseminated or extrapulmonary

Mycobacterium tuberculosis of any site, pulmonary, disseminated or extra-pulmonary

Mycobacterium, other species or unidentified species, disseminated† or extra-pulmonary

Pneumocystis jirovecii pneumonia (PJP)

Pneumonia, recurrent

Progressive multifocal leukoencephalopathy

Salmonella septicaemia, recurrent

Toxoplasmosis of brain, onset at age >1 month

Wasting syndrome attributed to HIV

Appendix B: Survey

Final Survey: Part 1

CASCADE | Concerted Action on SeroConversion to AIDS and Death in Europe

SECTION 1: COHORT INFORMATION

1. What is the name of your Cohort?

2. Are you able to provide us with details of your Standard Operating Procedures or Data Specification documentation as an attachment, or direct us to a website where an English Language copy is available?

☐ Yes –please send an attachment with the completed questionnaire if applicable

☐ No

Please place the website address in the box if applicable

3. Approximately how many clinics contribute data to your cohort, which are then shared with CASCADE?

4. Do you or your collaborating hospitals/clinics collect any information on Serious Non-AIDS events (SNAEs) from patients enrolled in CASCADE?

☐ Yes we do

☐ We do not collect any information but some/all of our collaborating clinics do (go to **Section 2**)

☐ No, neither we nor our collaborating clinics collect this information (go to **Section 3** at the end of the survey)

5. Do you contribute any data on Serious Non-AIDS events to the D:A:D study?

☐ Yes

☐ No (go to **Question 8**)

6. Approximately what percentage of CASCADE participants are enrolled in the D:A:D study?

0%

7. Are you currently completing D:A:D standardized case report forms for SNAEs for any CASCADE patients not enrolled in D:A:D?

☐ All Patients

☐ Some Patients – if so what percentage of your CASCADE patients are eligible to contribute?

Don't know %

☐ No Patients

8. Has your Cohort established Standardised Case Definitions for <u>any</u> SNAEs?

☐ Yes

☐ No (go to **Question 11**)

9. For which of the following SNAEs have you established Standardised Case Definitions?

☐	Myocardial Infarction	☐	Non-AIDS defining Cancers
☐	Congestive Heart Failure	☐	Fractures
☐	Diabetes Mellitus	☐	Pulmonary Embolism
☐	Coronary Revascularization	☐	Peripheral Arterial Disease
☐	End Stage Liver Disease	☐	Stroke
☐	End-Stage Renal Disease	☐	Deep Vein Thrombosis
☐	Acute Pancreatitis	☐	Other(s) – please specify

10. Is there any central adjudication of these events through an end point review committee?

☐ Yes for all events for which we have established Case Definitions

☐ For some events for which we have established Case Definitions

☐ No

11. Do you routinely collect information on any of the following SNAEs for patients enrolled in CASCADE?

☐	Myocardial Infarction	☐	Non-AIDS defining Cancers
☐	Congestive Heart Failure	☐	Fractures
☐	Diabetes Mellitus	☐	Pulmonary Embolism
☐	Coronary Revascularization	☐	Peripheral Arterial Disease
☐	End Stage Liver Disease	☐	Stroke
☐	End-Stage Renal Disease	☐	Deep Vein Thrombosis
☐	Acute Pancreatitis	☐	Other(s) – please specify

12. Do you use any of the following data sources to routinely collect or cross-check information on SNAEs for patients enrolled in CASCADE?

Yes	No	
☐	☐	Electronic health records (EHR) or Electronic medical records (EMR)
☐	☐	Registries e.g. Cancer registry
☐	☐	Electronic datasets e.g. clinics/hospitals send you datasets periodically
☐	☐	Detailed standardized paper case report forms for each event
☐	☐	Detailed standardized electronic case report forms for each event
☐	☐	Basic patient follow-up questionnaires completed on a regular basis

☐ ☐ Historical data on previous SNAEs in written or electronic format

☐ ☐ National death register

☐ ☐ Post-mortem/pathology reports

☐ ☐ Laboratory databases

☐ ☐ Other – please explain the information collection process below -

13. Please briefly describe the process by which you collect information using the data sources above.

14. Do you record your data for these events in a coded format?

☐ Yes

☐ Some aspects of the data are recorded in this way, but other details are recorded as free text

☐ No

15. Do you perform any of the following on any of your data?

Yes No

☐ ☐ Automated data cleaning at the time of data entry

☐ ☐ Manual data cleaning

☐ ☐ Query report generation i.e. your staff contact the clinics or use other data sources to attempt to clarify data discrepancies

☐ ☐ Data tracking i.e. methods are in place to identify problems by comparing indicators e.g. observed and expected rates of case report form completion are compared

☐ ☐ Auditing e.g. a random sample of patients' data are examined/verified

16. During data cleaning, do you routinely check the following?

Yes	No	Don't Know	
☐	☐	☐	Out of range values
☐	☐	☐	Missing values
☐	☐	☐	Duplicate records
☐	☐	☐	Invalid entries e.g. numerals in text field
☐	☐	☐	Logically inconsistent responses e.g. A female patient with prostate cancer
☐	☐	☐	That all negative HIV tests precede the first positive HIV test
☐	☐	☐	Sudden unexpected increases or decreases in CD4 Count/Viral Load
☐	☐	☐	Sudden falls in viral load before cART initiation

17. During cleaning of the SNAE data, do you routinely check the following?

Yes	No	Don't Know	
☐	☐	☐	Out of range values
☐	☐	☐	Missing values
☐	☐	☐	Invalid entries e.g. numerals in text field
☐	☐	☐	Logically inconsistent responses e.g. A female patient with prostate cancer

18. How systematically do your hospitals/clinics collect SNAE cases?

All/The Clinic	Some Clinics	No Clinics	
☐	☐	☐	Systematically for all patients from a given point in time

☐ ☐ ☐ Systematically for all patients admitted to the hospital from a given point in time, but not from those attending outpatient clinics

☐ ☐ ☐ Systematically for other patient groups from a given point in time – please specify for which groups

☐ ☐ ☐ On an ad-hoc basis for some patients

☐ ☐ ☐ Don't Know

19. Are your hospitals/clinics able to systematically record any SNAEs that have been diagnosed or treated at other hospitals or clinics?

☐ No

☐ Yes – If so please briefly describe how you capture these data and for which SNAEs

20. Do you systematically collect data on patients' smoking habits?

☐ Yes

☐ No (go to **Question 23**)

21. Do you systematically collect any of the following data do on patients' smoking habits?

Yes	No	
☐	☐	Ever Smoked
☐	☐	Current Smoker
☐	☐	Smoking status at each follow-up visit
☐	☐	Amount smoked – if so, briefly describe how this is recorded

22. For which patients do you record these data?

☐ All Patients from a given point in time

☐ Certain patient groups from a given point in time – please specify which groups

23. Do you systematically collect any data on patients' lipid levels?

☐ Yes

☐ No (go to **Question 28**)

24. Which of the following lipid data do you routinely collect?

☐ Total Cholesterol (TC)

☐ High density Lipoproteins (HDL) Cholesterol

☐ Low density Lipoproteins (LDL) Cholesterol – measured rather than derived from HDL/TC/TT

☐ Triglycerides

25. For which patients do you record these data?

☐ All Patients from a given point in time

☐ Certain patient groups from a given point in time – please specify which groups

26. How frequently/under what circumstances do you record these measurements?

☐ On enrolment to the study

☐ At each clinic visit

☐ At regular intervals during follow-up dependant on patients' previous lipid levels

☐ At regular intervals during follow-up independent of patients' previous lipid levels

☐ Other – please specify

27. Have you systematically undertaken any historical data collection of lipid levels?

☐ Yes

☐ No

28. Do you systematically collect data on patients' weight?

☐ Yes

☐ No (go to **Question 31**)

29. For which patients do you record these data?

☐ All Patients from a given point in time

☐ Certain patient groups from a given point in time – please specify which groups

30. How frequently/under what circumstances do you record patients' weight?

☐ On enrolment to the study

☐ At each clinic visit

☐ At least annually during follow-up

☐ At regular intervals dependent on patients' previous weight/BMI

☐ Other – please specify

31. Do you collect data on patients' blood pressure?

☐ Yes

☐ No (go to **Question 34**)

32. For which patients do you record these data?

☐ All Patients from a given point in time

☐ Certain patient groups from a given point in time – please specify which groups

33. How frequently/under what circumstances do you record these measurements?

☐ On enrolment to the study

☐ At each clinic visit

☐ At regular intervals during follow-up dependant on patients' previous blood pressure

☐ At least annually during follow-up

☐ Other – please specify

34. Do you systematically measure patients' height?

☐ Yes

☐ No (go to **Question 37**)

35. For which patients do you record these data?

☐ All Patients from a given point in time

☐ Certain patient groups from a given point in time – please specify which groups

36. Under what circumstances do you record these measurements?

☐ On enrolment to the study

☐ Other – please specify

37. Do you systematically collect data on patients' drinking habits?

☐ Yes

☐ No (go to **Question 40**)

38. Do you systematically collect any of the following data do on patients' alcohol consumption?

Yes	No	
☐	☐	History of alcohol problem
☐	☐	Current alcohol problem
☐	☐	Alcohol consumption discussion at each follow-up visit
☐	☐	Amount consumed – if so, briefly describe how this is recorded

39. For which patients do you record these data?

☐ All Patients from a given point in time

☐ Certain patient groups from a given point in time – please specify which groups

40. What statistical analyses, if any, has your Cohort undertaken involving SNAEs?

☐ We have published analyses involving SNAEs in peer reviewed journals

☐ We have undertaken analyses involving SNAEs and plan to publish it in the near future

☐ We have undertaken analyses involving SNAEs but have no plans to publish in the near future

☐ We have not undertaken any analyses involving SNAEs but plan to do so in the near future

☐ We have not undertaken any analyses involving SNAEs and currently have no plans to do so

41. Are the following events currently being <u>systematically</u> recorded by any of your <u>clinics</u>? This may or may not differ from what you yourself are recording. Please also indicate for approximately what percentage of the <u>patients enrolled in CASCADE</u> this information is currently being recorded, if known. If you are contributing data from a single clinic please select 100%.

Yes	No	Don't Know		
☐	☐	☐	Myocardial Infarction	Don't know %
☐	☐	☐	Congestive Heart Failure	Don't know %

☐	☐	☐	Diabetes Mellitus	Don't know %
☐	☐	☐	Coronary Revascularization	Don't know %
☐	☐	☐	Decompensated Liver Disease	Don't know %
☐	☐	☐	End-Stage Renal Disease	Don't know %
☐	☐	☐	Non-AIDS defining Cancers	Don't know %
☐	☐	☐	Fractures	Don't know %
☐	☐	☐	Pulmonary Embolism	Don't know %
☐	☐	☐	Peripheral Arterial Disease	Don't know %
☐	☐	☐	Stroke	Don't know %
☐	☐	☐	Deep Vein Thrombosis	Don't know %
☐	☐	☐	Acute Pancreatitis	Don't know %

END

This is the end of the survey, thank you so much for taking the time to complete it. Please would you return it to me as an attachment and include any additional information which you think I might find useful.

SECTION 2: CLINIC INFORMATION

1. Do your hospitals/clinic use any of the following methods to systematically collect information on SNAEs for patients enrolled in CASCADE?

Yes All/The Clinic	Yes Some Clinics	No Clinics	
☐	☐	☐	Electronic health records (EHR) or Electronic medical records (EMR)
☐	☐	☐	Written clinical notes
☐	☐	☐	Detailed standardized paper case report forms for each event
☐	☐	☐	Detailed standardized electronic case report forms for each event
☐	☐	☐	A hybrid system using both computerized and written clinical records
☐	☐	☐	Computerized database designed by the cohort
	☐		Other – please specify

2. From which patients do your hospitals/clinics identify SNAE cases?

Yes All/The Clinic	Yes Some Clinics	No Clinics	
☐	☐	☐	All patients from a given point in time
☐	☐	☐	Patients admitted to the hospital, but NOT from those attending outpatient Clinics from a given point in time
☐	☐	☐	Other patient groups from a given point in time – please specify which groups and when

3. Are your hospitals/clinics able to systematically record any SNAES that have been diagnosed or treated at other hospitals or clinics not run by them?

☐ No

☐ Yes – If so please briefly describe how you capture these data and for which SNAEs

4. Are the disease coding systems used to record data on SNAEs the same across all your clinics? (If there is just one clinic please just choose between the first two options below)

☐ Yes –ICD-9/ICD-10 codes are used

☐ Yes –Other standardized codes are used e.g. codes created by the cohort or other national or regional coding systems (e.g. SNOMED/MedDRA) please specify

☐ No – some/all clinics use different coding systems but we have mapped them so that the data we have collected are directly comparable (e.g. local codes have all been converted to ICD-9)

☐ No – they use different codes and we have not undertaken any mapping so the data we have collected from different clinics are not currently directly comparable

5. If your clinic(s) currently use different coding systems for SNAEs, are any of the following actions likely to be able to be undertaken by you or your clinics in the future?

Yes	No	
☐	☐	Standardization of coding across clinics
☐	☐	Mapping of codes between clinics to create comparable diagnoses across clinics
☐	☐	The creation of Standardized case report forms on which clinics can report SNAEs
☐		Not applicable

6. Are the following events currently being systematically recorded by any of your clinics? This may or may not differ from what you yourself are recording. Please also indicate for approximately what percentage of the patients enrolled in CASCADE this information is currently being recorded, if known. If you are contributing data from a single clinic please select 100%.

Yes	No	Don't Know		
☐	☐	☐	Myocardial Infarction	Don't know %
☐	☐	☐	Congestive Heart Failure	Don't know %
☐	☐	☐	Diabetes Mellitus	Don't know %

☐	☐	☐	Coronary Revascularization	Don't know %
☐	☐	☐	Decompensated Liver Disease	Don't know %
☐	☐	☐	End-Stage Renal Disease	Don't know %
☐	☐	☐	Non-AIDS defining Cancers	Don't know %
☐	☐	☐	Fractures	Don't know %
☐	☐	☐	Pulmonary Embolism	Don't know %
☐	☐	☐	Peripheral Arterial Disease	Don't know %
☐	☐	☐	Stroke	Don't know %
☐	☐	☐	Deep Vein Thrombosis	Don't know %
☐	☐	☐	Acute Pancreatitis	Don't know %

END

This is the end of the survey, thank you so much for taking the time to complete it. Please would you return it to me as an attachment and include any additional information which you think I might find useful

SECTION 3: FUTURE DATA COLLECTION

1. Do you plan to start collecting data on SNAEs in the near future?

☐ Yes What and When? (go to **END**)

☐ No Why not?

2. Do any of your collaborating clinics plan to start collecting any data on Serious Non-AIDS events in the near future?

☐ Yes What and When? (go to **END**)

☐ No

☐ Do not know - would it be possible to find out?

3. Do you think any of your collaborating clinics have the resources to collect this information?

☐ Yes

☐ No

☐ Do not know – would it be possible to find out?

END

This is the end of the survey, thank you so much for taking the time to complete it. Please would you return it to me as an attachment and include any additional information which you think I might find useful.

MYOCARDIAL INFARCTION

SECTION 1

1. Do you collect ANY data on Myocardial Infarctions?

☐ Yes (Go to **Question 3**)

☐ No

2. Do your collaborating clinics collect ANY data on Myocardial Infarctions?

☐ Yes (Go to **Question 12**)

☐ No (Go to **Section 2**)

☐ Don't Know- Would it be possible to find out? (Go to **Section 2**)

3. When did you start systematically collecting data on Myocardial Infarctions in a prospective manner?

4. Approximately how many Events have been recorded (if known)?

5. Approximately how many patient years of follow up have been recorded (if known)?

6. Have you systematically collected any retrospective data on these events?

☐ Yes – if yes what, when and how have you been collecting it

☐ No

7. Has your Cohort determined a Standardized Case Definition(s) for Myocardial Infarction?

☐ Yes

☐ No (if no go to **Question 10**)

8. Which of the following form part of your Case Definition(s)?

☐ Troponin levels (T or I)

☐ CK-MB levels

☐ CK levels

☐ LDH levels (1 and 2)

☐ ECG: Development of pathologic Q waves

☐ ECG: Development of new ST-T changes

☐ Ischemic symptoms (e.g. Chest Pain)

☐ Other – please specify

☐ Imaging evidence of new loss of viable myocardium

☐ Imaging evidence of new regional wall motion abnormality

☐ Sudden unexpected cardiac death

☐ Pathological findings of acute myocardial infarction

☐ Pathological findings of a healed or healing myocardial infarction

☐ Diagnostic codes (e.g. ICD-9/ICD-10)

☐ Autopsy report of Myocardial Infarction

9. Is there any central adjudication of these events through an end point review committee?

☐ Yes

☐ No

10. How comprehensively are you recording these Myocardial events?

☐ These events have been collected systematically for all/most patients from a given point in time

☐ These events have been collected systematically for some patients from a given point in time

☐ These data has been collected but not systematically and/or recording is erratic

11. What data are you recording relating to these events for each patient?

Yes	No	
☐	☐	Event date
☐	☐	Clinical Findings
☐	☐	Serology results e.g. Tropin/CK/LDH levels
☐	☐	ECG results
☐	☐	Imaging findings e.g. Ultrasound
☐	☐	ICD-9/ICD-10 codes
☐	☐	Autopsy findings
☐	☐	Case definition based splitting of events into confirmed and probable categories
☐	☐	History of previous cardiovascular events
☐	☐	Information on whether this diagnosis has been made elsewhere
☐	☐	Historical and/or current cardiovascular drug treatment
☐	☐	Smoking history
☐	☐	Weight

12. How completely have your collaborating clinics/hospitals recorded data on the occurrence of Myocardial Infarctions – this may or may not differ from what you yourselves are recording?

☐ These events have been collected systematically for all/most patients from a given point in time

☐ These events have been collected systematically for some patients from a given point in time

☐ These events have been collected but not systematically and/or recording is erratic

☐ We know that at least some clinics collect this data but we do have details of when they started and/or whether it is collected and recorded systematically.

13. What information are your collaborating clinics currently collecting relating to Myocardial Infarctions?

All Clinics	Some Clinics	No Clinics	
☐	☐	☐	Event date
☐	☐	☐	Clinical Findings
☐	☐	☐	Serology results e.g. Tropin/CK/LDH levels
☐	☐	☐	ECG results
☐	☐	☐	Imaging findings e.g. Ultrasound
☐	☐	☐	ICD-9/ICD-10 codes
☐	☐	☐	Autopsy findings
☐	☐	☐	History of previous cardiovascular events
☐	☐	☐	Information on whether this diagnosis has been made elsewhere
☐	☐	☐	Historical and/or current cardiovascular drug treatment
☐	☐	☐	Smoking history
☐	☐	☐	Weight
☐	☐	☐	We do not know what they are collecting (would it be possible to find out?)

CONGESTIVE HEART FAILURE

SECTION 2

1. Do you collect ANY data on Congestive Heart Failure?

☐ Yes (Go to **Question 3**)

☐ No

2. Do your collaborating clinics collect ANY data on Congestive Heart Failure?

☐ Yes (Go to **Question 12**)

☐ No (Go to **Section 3**)

☐ Don't Know- Would it be possible to find out? (Go to **Section 3**)

3. When did you start systematically collecting data on Congestive Heart Failure in a prospective manner?

4. Approximately how many Events have been recorded (if known)?

5. Approximately how many patient years of follow up have been recorded (if known)?

6. Have you systematically collected any retrospective data on these events?

☐ Yes – if yes what, when and how have you been collecting it

☐ No

7. Has your Cohort determined a Standardized Case Definition(s) for Congestive Heart Failure?

☐ Yes

☐ No (if no go to **Question 10**)

8. Please put your Case Definition in the text box below?

9. Is there any central adjudication of these events through an end point review committee?

☐ Yes

☐ No

10. How comprehensively are you recording these Myocardial events?

☐ These events have been collected systematically for all/most patients from a given point in time

☐ These events have been collected systematically for some patients from a given point in time

☐ These data has been collected but not systematically and/or recording is erratic

11. What data are you recording relating to these events for each patient?

Yes	No	
☐	☐	Event date
☐	☐	Clinical Findings
☐	☐	BNP/NTproBNP results
☐	☐	ECG results
☐	☐	Radiographic findings
☐	☐	Ejection fraction
☐	☐	ICD-9/ICD-10 codes
☐	☐	Autopsy findings
☐	☐	Case definition based splitting of events into confirmed and probable categories
☐	☐	History of previous cardiovascular events
☐	☐	Information on whether this diagnosis has been made elsewhere
☐	☐	Historical and/or current cardiovascular drug treatment

☐ ☐ Smoking history

☐ ☐ Blood Pressure

☐ ☐ Weight

12. How completely have your collaborating clinics/hospitals recorded data on the occurrence of Congestive Heart Failure – this may or may not differ from what you yourselves are recording?

☐ These events have been collected systematically for all/most patients from a given point in time

☐ These events have been collected systematically for some patients from a given point in time

☐ These events have been collected but not systematically and/or recording is erratic

☐ We know that at least some clinics collect this data but we do have details of when they started and/or whether it is collected and recorded systematically.

13. What information are your collaborating clinics currently collecting relating to Congestive Heart Failure?

All Clinics	Some Clinics	No Clinics	
☐	☐	☐	Event date
☐	☐	☐	Clinical Findings
☐	☐	☐	BNP/NTproBNP
☐	☐	☐	ECG results
☐	☐	☐	Radiographic findings
☐	☐	☐	Ejection fraction
☐	☐	☐	ICD-9/ICD-10 codes
☐	☐	☐	Autopsy findings
☐	☐	☐	History of previous cardiovascular events
☐	☐	☐	Information on whether this diagnosis has been made elsewhere

☐ ☐ ☐ Historical and/or current cardiovascular drug treatment

☐ ☐ ☐ Smoking history

☐ ☐ ☐ Weight

☐ ☐ ☐ Blood Pressure

☐ ☐ ☐ We do not know what they are collecting (would it be possible to find out?)

DIABETES MELLITUS (Type I and Type II)

SECTION 3

1. Do you collect ANY data on Diabetes Mellitus?

☐ Yes (Go to **Question 3**)

☐ No

2. Do your collaborating clinics collect ANY data on Diabetes Mellitus?

☐ Yes (Go to **Question 11**)

☐ No (Go to **Section 4**)

☐ Don't Know – Would it be possible to find out? (Go to **Section 4**)

3. When did you start systematically collecting data on Diabetes Mellitus in a prospective manner?

4. Approximately how many Events have been recorded (if known)?

Approximately how many patient years of follow up have been recorded (if known)?

5. Have you systematically collected any retrospective data on these events?

☐ Yes

☐ No

6. Has your Cohort determined a Standardized Case Definition for Diabetes Mellitus?

☐ Yes

☐ No (if no go to **Question 9**)

7. Which of the following form part of your Case Definition?

☐ Polyuria and Polydypsia

☐ Glycated Haemoglobin (HbA1c)

☐ Plasma glucose concentration (taken at any time) >11.1mmol/L (>200mg/dL)

☐ Fasting plasma glucose concentration >7.0 mmol/L (>126 mg/dL)

☐ Oral glucose tolerance test >11.1mmol/L (>200mg/dL)

☐ Diagnostic codes (ICD-9/ICD-10 etc)

☐ Other – please specify

8. Is there any central adjudication of these events through an end point review committee?

☐ Yes

☐ No

9. How comprehensively are you recording Diabetes Mellitus diagnoses?

☐ These diagnoses have been collected systematically for all/most patients from a given point in time

☐ These diagnoses have been collected systematically for some patients from a given point in time

☐ These diagnoses have been collected but not systematically and/or recording is erratic

10. What data are you recording relating to these events for each patient?

Yes	No	
☐	☐	Diagnosis date
☐	☐	Plasma Glucose Concentrations
☐	☐	Glycated Haemoglobin (HbA1C)
☐	☐	Oral glucose tolerance test results

☐ ☐ History of previous pancreatitis

☐ ☐ ICD-9/ICD-10 codes

☐ ☐ History of medical treatment other than ART that might have precipitated diabetes

☐ ☐ Information on whether this diagnosis was made elsewhere

11. How completely have your collaborating clinics/hospitals recorded data on the diagnosis of Diabetes Mellitus– this may or may not differ from what you yourselves are recording?

☐ These diagnoses have been collected systematically for all/most patients from a given point in time

☐ These diagnoses have been collected systematically for some patients from a given point in time

☐ These diagnoses have been collected but not systematically and/or recording is erratic

☐ We know that at least some clinics collect these diagnoses but we do have details of when they started and/or whether it is collected and recorded systematically.

12. What information are your collaborating clinics currently collecting relating to Diabetes Mellitus?

All Clinics	Some Clinics	No Clinics	
☐	☐	☐	Diagnosis date
☐	☐	☐	Plasma Glucose Concentrations
☐	☐	☐	Glycated Haemoglobin (HbA1c)
☐	☐	☐	Oral glucose tolerance test results
☐	☐	☐	History of previous pancreatitis
☐	☐	☐	ICD-9/ICD-10 codes
☐	☐	☐	Non-ART medical treatment history that might have precipitated diabetes
☐	☐	☐	Information on whether this diagnosis was made elsewhere
☐	☐	☐	We do not know what they are collecting (would it be possible to find out?)

CORONARY REVASCULARIZATION

SECTION 4

1. Do you collect ANY data on Coronary Revascularization (coronary artery by-pass grafting/coronary angioplasty/coronary stenting/coronary endarterectomy)?

☐ Yes (Go to **Question 3**)

☐ No

2. Do your collaborating clinics collect ANY data on Coronary Revascularization?

☐ Yes (Go to **Question 10**)

☐ No (Go to **Section 5**)

☐ Don't Know -Would it be possible to find out? (Go to **Section 5**)

3. When did you start systematically collecting data on Coronary Revascularization in a prospective manner?

4. Approximately how many Events have been recorded (if known)?

5. Approximately how many patient years of follow up have been recorded (if known)?

6. Have you systematically collected any retrospective data on these events?

☐ Yes

☐ No

7. On which events do you collect information?

☐ Coronary Artery by-pass grafting

☐ Coronary Angioplasty

☐ Coronary Stenting

☐ Coronary Endarterectomy

8. How comprehensively are you recording Coronary Revascularization procedures?

☐ These procedures have been collected systematically for all/most patients from a given point in time

☐ These procedures have been collected systematically for some patients from a given point in time

☐ These procedures have been collected but not systematically and/or recording is erratic

9. What data are you recording relating to these procedures for each patient?

Yes	No	
☐	☐	Procedure Date
☐	☐	Type of procedure performed
☐	☐	Whether the procedure was conducted in relation to a Myocardial Infarction
☐	☐	Whether the procedure was complicated by a Stroke
☐	☐	Whether the procedure was performed elsewhere

10. How completely have your collaborating clinics/hospitals recorded data on Coronary Revascularization procedures– this may or may not differ from what you yourselves are recording?

☐ These procedures have been collected systematically for all/most patients from a given point in time

☐ These procedures have been collected systematically for some patients from a given point in time

☐ These procedures have been collected but not systematically and/or recording is erratic

☐ We know that at least some clinics collect these procedures but we do have details of when they started and/or whether it is collected and recorded systematically.

11. What information are your collaborating clinics currently collecting relating to Coronary Revascularization?

All Some No

Clinics	Clinics	Clinics	
☐	☐	☐	Procedure Date
☐	☐	☐	Whether the procedure was conducted in relation to a Myocardial Infarction
☐	☐	☐	Whether the procedure was complicated by a Stroke
☐	☐	☐	We do not know what they are collecting (would it be possible to find out?)

END-STAGE LIVER DISEASE

SECTION 5

1. Do you collect ANY data on End Stage Liver Disease (ESLD)?

☐ Yes (Go to **Question 3**)

☐ No

2. Do your collaborating clinics collect ANY data on Decompensated Liver Failure?

☐ Yes (Go to **Question 12**)

☐ No (**Go to Section 6**)

☐ Don't Know – if possible could you find out? (Go to **Section 6**)

3. When did you start systematically collecting data on End Stage Liver Disease in a prospective manner?

4. Approximately how many Events have been recorded (if known)?

5. Approximately how many patient years of follow up have been recorded (if known)?

6. Have you systematically collected any retrospective data on these events?

☐ Yes

☐ No

7. Has your Cohort determined a Standardized Case Definition for ESLD?

☐ Yes

☐ No (if no go to **Question 10**)

8. Which of the following form part of your Case Definition?

☐ Evidence of cirrhosis on histological samples obtained from autopsy or biopsy

☐ Evidence of cirrhosis on MRI or CT scan

☐ Ultrasound imaging consistent with cirrhosis

☐ Hepato-Renal syndrome

☐ Ascites without an alternative explanation

☐ Hepatic encephalopathy

☐ Bleeding from gastric or oesophageal varices

☐ Clinical evidence of spontaneous bacterial peritonitis

☐ Liver Transplantation

☐ Diagnostic codes (e.g. ICD-9/ICD-10)

☐ High MELD/UKELD scores (Model for End-Stage Liver disease)

☐ Other – please specifiy

9. Is there any central adjudication of these events through an end point review committee?

☐ Yes

☐ No

10. How comprehensively are you recording End Stage Liver Disease diagnoses?

☐ These diagnoses have been collected systematically for all/most patients from a given point in time

☐ These diagnoses have been collected systematically for some patients from a given point in time

☐ These diagnoses have been collected but not systematically and/or recording is erratic

11. What data are you recording relating to these events for each patient?

Yes	No	
☐	☐	Diagnosis Date
☐	☐	ICD-9/ICD-10 Codes
☐	☐	Liver biopsy and/or Fibroscan and/or other imaging results
☐	☐	Case definition based splitting of events into confirmed and probable categories
☐	☐	History of previous Liver Disease
☐	☐	Hepatitis B infection
☐	☐	Hepatitis C infection
☐	☐	Current or past alcoholism

12. How completely have your collaborating clinics/hospitals recorded data on the diagnosis of End Stage Liver Disease– this may or may not differ from what you yourselves are recording?

☐ These diagnoses have been collected systematically for all/most patients from a given point in time

☐ These diagnoses have been collected systematically for some patients from a given point in time

☐ These diagnoses have been collected but not systematically and/or recording is erratic

☐ We know that at least some clinics collect these diagnoses but we do have details of when they started and/or whether it is collected and recorded systematically.

13. What information are your collaborating clinics currently collecting relating to ESLD?

All Clinics	Some Clinics	No Clinics	
☐	☐	☐	Diagnosis Date
☐	☐	☐	ICD-9/ICD-10 Codes

☐ ☐ ☐ Liver biopsy and/or Fibroscan and/or other imaging results

☐ ☐ ☐ History of previous Liver Disease

☐ ☐ ☐ Hepatitis B infection

☐ ☐ ☐ Hepatitis C infection

☐ ☐ ☐ Current or past alcoholism

☐ ☐ ☐ History of medication other than ART that might have precipitated Liver Failure

☐ ☐ ☐ We do not know what they are collecting (would it be possible to find out?)

END-STAGE RENAL DISEASE

SECTION 6

1. Do you collect ANY data on End-Stage Renal Disease?

☐ Yes (Go to **Question 3**)

☐ No

2. Do your collaborating clinics collect ANY data on End-Stage Renal Disease?

☐ Yes (Go to **Question 11**)

☐ No (Go **to <u>Section</u> 7**)

☐ Don't Know – would it be possible to find out? (Go to **<u>Section</u> 7**)

3. When did you start collecting data on End-Stage Renal Disease?

4. Approximately how many Events have been recorded (if known)?

5. Approximately how many patient years of follow up have been recorded (if known)?

6. Has your Cohort determined a Standardized Case Definition for End-Stage Renal Disease?

☐ Yes

☐ No (if no go to **Question 9**)

7. Which of the following form part of your Case Definition?

☐ Kidney transplantation

☐ Haemodialysis for more than 3 months

☐ Peritoneal dialysis for more than 3 months

☐ Haemodialysis or peritoneal dialysis for more than 1 month and up until the time of the patient's death, for a patient who dies before 3 months

☐ GFR < given value e.g. 15ml/min/1.73m^2

☐ Diagnostic codes (e.g. ICD-9/ICD-10)

☐ Other – please specify

8. Is there any central adjudication of these events through an end point review committee?

☐ Yes

☐ No

9. How comprehensively are you recording End-Stage Renal Disease diagnoses?

☐ These diagnoses have been collected systematically for all/most patients from a given point in time

☐ These diagnoses have been collected systematically for some patients from a given point in time

☐ These diagnoses have been collected but not systematically and/or recording is erratic

10. What data are you recording relating to these events for each patient?

Yes	No	
☐	☐	Diagnosis Date
☐	☐	Clinical Findings
☐	☐	ICD-9/ICD-10 Codes
☐	☐	Creatinine
☐	☐	Urea
☐	☐	Electrolytes
☐	☐	GFR
☐	☐	Case definition based splitting of events into <u>confirmed</u> and <u>probable</u> categories

☐ ☐ History of previous Renal Disease

☐ ☐ Histology results

☐ ☐ Ethnicity

☐ ☐ History of medication other than ART that might have precipitated renal failure

11. How completely have your collaborating clinics/hospitals recorded data on the diagnosis of End Stage Renal Disease – this may or may not differ from what you yourselves are recording?

☐ These diagnoses have been collected systematically for all/most patients from a given point in time

☐ These diagnoses have been collected systematically for some patients from a given point in time

☐ These diagnoses have been collected but not systematically and/or recording is erratic

☐ We know that at least some clinics collect these diagnoses but we do have details of when they started and/or whether it is collected and recorded systematically.

12. What information are your collaborators currently collecting relating to End Stage Renal Disease?

All Clinics	Some Clinics	No Clinics	
☐	☐	☐	Diagnosis Date
☐	☐	☐	ICD-9/ICD-10 Codes
☐	☐	☐	Creatinine
☐	☐	☐	Urea
☐	☐	☐	Electrolytes
☐	☐	☐	GFR
☐	☐	☐	Ethnicity
☐	☐	☐	History of previous Renal Disease
☐	☐	☐	Histology results

☐ ☐ ☐ Records of Medication other than ART that might have caused renal failure

☐ ☐ ☐ We do not know what they are collecting (would it be possible to find out?)

13. If collaborating clinics are collecting GFR what methods do are they using to calculate this?

All Clinics	Some Clinics	No Clinics	
☐	☐	☐	Cockcroft-Gault formula
☐	☐	☐	MDRD formula
☐	☐	☐	CKD-Epi formula
☐	☐	☐	May Quadratic formula
☐	☐	☐	Inulin clearance
☐	☐	☐	Creatinine clearance
☐	☐	☐	We do not know how they are measuring it

NON-AIDS DEFINING CANCERS

SECTION 7

1. Do you collect ANY data on Non-Aids defining cancers (i.e. not Kaposi's Sarcoma (KS), invasive cervical cancer or non-Hodgkin Lymphoma)?

☐ Yes (Go to **Question 3**)

☐ No

2. Do your collaborating clinics collect ANY data on Non-Aids Defining Cancers?

☐ Yes (Go to **Question 12**)

☐ No (Go to **Section 8**)

☐ Don't Know- would it be possible to find out? (Go to **Section 8**)

3. When did you start collecting data on Non-Aids Defining Cancers?

4. Approximately how many Events have been recorded (if known)?

5. Approximately how many patient years of follow up have been recorded (if known)?

6. Which of the following cancers are you collecting data on?

☐ Lung/Bronchus	☐ Anal	☐ Hodgkin's Lymphoma
☐ Kidney	☐ Liver	☐ Head and neck
☐ Multiple Myeloma	☐ Vulva/Vagina	☐ Melanoma
☐ Prostate	☐ Brain/CNS	☐ Non-Melanoma Skin
☐ Penis	☐ All cancers	☐ Leukaemia

Other, please specify

7. Has your Cohort determined a Standardized Case Definition(s) for any Non-Aids Defining Cancers?

☐ Yes

☐ No (if no go to **Question 10**)

8. Please give full details of your Case Definitions if possible (please send as an attachment if needed):-

9. Is there any central adjudication of these events through an end point review committee?

☐ Yes

☐ No

10. How comprehensively are you recording non-Aids defining cancer diagnoses?

☐ These diagnoses have been collected systematically for all/most patients from a given point in time

☐ These diagnoses have been collected systematically for some patients from a given point in time

☐ These diagnoses have been collected but not systematically and/or recording is erratic

11. What data are you recording relating to these events for each patient?

Yes	No	
☐	☐	Diagnosis Date
☐	☐	ICD-9/ICD-10 Codes
☐	☐	Stage at diagnosis
☐	☐	Imaging results
☐	☐	Histology/cytology results

☐ ☐ Biochemical assays e.g. cancer cell markers, alpha-fetoprotein

☐ ☐ Case definition based splitting of events into confirmed and probable categories

☐ ☐ History of previous tumours/radiotherapy/chemotherapy

☐ ☐ Information on whether this diagnosis was made elsewhere

12. How completely have your collaborating clinics/hospitals recorded data on the diagnosis of non-Aids defining Cancers– this may or may not differ from what you yourselves are recording?

☐ These diagnoses have been collected systematically for all/most patients from a given point in time

☐ These diagnoses have been collected systematically for some patients from a given point in time

☐ These diagnoses have been collected but not systematically and/or recording is erratic

☐ We know that at least some clinics collect these diagnoses but we do have details of when they started and/or whether it is collected and recorded systematically.

13. What information are your collaborating clinics currently collecting relating to non-Aids defining Cancer?

All Clinics	Some Clinics	No Clinics	
☐	☐	☐	Diagnosis Date
☐	☐	☐	Clinical Findings
☐	☐	☐	ICD-9/ICD-10 Codes
☐	☐	☐	Stage at diagnosis
☐	☐	☐	Imaging results
☐	☐	☐	Histology/cytology results
☐	☐	☐	Biochemical assays e.g. cancer cell markers, alpha-fetoprotein
☐	☐	☐	History of previous tumours/radiotherapy/chemotherapy
☐	☐	☐	We do not know what they are collecting (would it be possible to find out?)

FRACTURES

SECTION 8

1. Do you collect ANY data on Fractures?

☐ Yes (Go to **Question 3**)

☐ No

2. Do your collaborating clinics collect ANY data on Fractures?

☐ Yes (Go to **Question 9**)

☐ No (Go to **Section 9**)

☐ Don't Know- would it be possible to find out? (Go to **Section 9**)

3. When did you start collecting data on Fractures?

4. Approximately how many Events have been recorded (if known)?

5. Approximately how many patient years of follow up have been recorded (if known)?

6. Which of the following types of fractures do you collect data on?

☐ Traumatic Fractures

☐ Pathological Fractures

☐ Stress Fractures

7. How comprehensively are you recording fracture diagnoses?

☐ These diagnoses have been collected systematically for all/most patients from a given point in time

☐ These diagnoses have been collected systematically for some patients from a given point in time

☐ These diagnoses have been collected but not systematically and/or recording is erratic

8. What data are you recording relating to these events for each patient?

Yes	No	
☐	☐	Diagnosis Date
☐	☐	ICD-9/ICD-10 Codes
☐	☐	Fracture type i.e. Traumatic/Pathological/Stress
☐	☐	Fracture location
☐	☐	Imaging results
☐	☐	History of underlying disease which might predispose to fractures e.g. osteoporosis
☐	☐	Whether this diagnosis was made elsewhere

9. How completely have your collaborating clinics/hospitals recorded data on Fractures– this may or may not differ from what you yourselves are recording?

☐ These diagnoses have been collected systematically for all/most patients from a given point in time

☐ These diagnoses have been collected systematically for some patients from a given point in time

☐ These diagnoses have been collected but not systematically and/or recording is erratic

☐ We know that at least some clinics collect these diagnoses but we do have details of when they started and/or whether it is collected and recorded systematically.

10. What information are your collaborating clinics currently collecting relating to Fractures?

All Clinics	Some Clinics	No Clinics	
☐	☐	☐	Diagnosis Date
☐	☐	☐	ICD-9/ICD-10 Codes
☐	☐	☐	Fracture type i.e. Traumatic/Pathological/Stress
☐	☐	☐	Fracture location
☐	☐	☐	Imaging results
☐	☐	☐	History of underlying disease which might predispose to fractures
☐	☐	☐	We do not know what they are collecting (would it be possible to find out?)

PULMONARY EMBOLISM

SECTION 9

1. Do you collect ANY data on Pulmonary Embolism?

☐ Yes (Go to **Question 3**)

☐ No

2. Do your collaborating clinics collect ANY data on Pulmonary Embolism?

☐ Yes (Go to **Question 11**)

☐ No (Go to **Section 10**)

☐ Don't Know- Would it be possible for you to find out? (Go to **Section 10**)

3. When did you start collecting data on Pulmonary Embolism?

4. Approximately how many Events have been recorded (if known)?

5. Approximately how many patient years of follow up have been recorded (if known)?

6. Has your Cohort determined a Standardized Case Definition for Pulmonary Embolism?

☐ Yes

☐ No (if no go to **Question 9**)

7. Which of the following form part of your Case Definition?

☐ Clinical signs compatible with Pulmonary Embolism such as chest pain, shortness of breath or haemoptysis

☐ Imaging compatible with Pulmonary Embolism such as pulmonary angiography, helical CT, ventilation-perfusion scan, ultrasound or venography

☐ Imaging studies compatible with a diagnosis of DVT

☐ Autopsy report of Pulmonary Embolism

☐ Elevated plasma D-dimer levels

☐ Diagnostic Codes (e.g. ICD-9/ICD-10)

☐ Other – please specify

8. Is there any central adjudication of these events through an end point review committee?

☐ Yes

☐ No

9. How comprehensively are you recording Pulmonary Embolism diagnoses?

☐ These diagnoses have been collected systematically for all/most patients from a given point in time

☐ These diagnoses have been collected systematically for some patients from a given point in time

☐ These diagnoses have been collected but not systematically and/or recording is erratic

10. What data are you recording relating to these events for each patient?

Yes	No	
☐	☐	Diagnosis Date
☐	☐	Clinical Findings
☐	☐	ICD-9/ICD-10 Codes
☐	☐	Autopsy results
☐	☐	Imaging results
☐	☐	Case definition based splitting of events into confirmed and probable categories

☐ ☐ D-dimer levels

☐ ☐ Whether this diagnosis was made elsewhere

11. How completely have your collaborating clinics/hospitals recorded data on the diagnosis of Pulmonary Embolism – this may or may not differ from what you yourselves are recording?

☐ These diagnoses have been collected systematically for all/most patients from a given point in time

☐ These diagnoses have been collected systematically for some patients from a given point in time

☐ These diagnoses have been collected but not systematically and/or recording is erratic

☐ We know that at least some clinics collect these diagnoses but we do have details of when they started and/or whether it is collected and recorded systematically.

12. What information are your collaborating clinics currently collecting relating to Pulmonary Embolism?

All Clinics	Some Clinics	No Clinics	
☐	☐	☐	Diagnosis Date
☐	☐	☐	ICD-9/ICD-10 Codes
☐	☐	☐	Autopsy results
☐	☐	☐	Imaging results
☐	☐	☐	D-dimer levels
☐	☐	☐	We do not know what they are collecting (would it be possible to find out?)

PERIPHERAL ARTERIAL DISEASE (PAD)

SECTION 10

1. Do you collect ANY data on Peripheral Arterial Disease?

☐ Yes (Go to **Question 3**)

☐ No

2. Do your collaborating clinics collect ANY data on Peripheral Arterial Disease?

☐ Yes (Go to **Question 11**)

☐ No (Go to **Section 11**)

☐ Don't Know- Would it be possible for you to find out? (Go to **Section 11**)

3. When did you start collecting data on Peripheral Arterial Disease?

4. Approximately how many Events have been recorded (if known)?

5. Approximately how many patient years of follow up have been recorded (if known)?

6. Has your Cohort determined a Standardized Case Definition for Peripheral Arterial Disease?

☐ Yes

☐ No (if no go to **Question 9**)

7. Which of the following form part of your Case Definition?

☐ Claudication

☐ Reduced hair and nail growth

☐ Minor wounds and sores which are slow to heal or do not so

☐ Colour and temperature changes in the limb

☐ Weak leg pulses on clinical examination

☐ Reduced Ankle-Brachial Index

☐ Evidence of PAD on Duplex Ultrasound

☐ Evidence of PAD on CT scan

☐ Evidence of PAD on Magnetic Resonance Angiography

☐ Evidence of PAD on Angiography

☐ Diagnostic codes (e.g. ICD-9/ICD-10)

☐ Other – please specify

8. Is there any central adjudication of these events through an end point review committee?

☐ Yes

☐ No

9. How comprehensively are you recording Peripheral Arterial Disease diagnoses?

☐ These diagnoses have been collected systematically for all/most patients from a given point in time

☐ These diagnoses have been collected systematically for some patients from a given point in time

☐ These diagnoses have been collected but not systematically and/or recording is erratic

10. What data are you recording relating to these events for each patient?

Yes	No	
☐	☐	Diagnosis Date
☐	☐	Clinical Findings
☐	☐	ICD-9/ICD-10 Codes

☐ ☐ Imaging results

☐ ☐ Case definition based splitting of events into confirmed and probable categories

☐ ☐ Whether the diagnosis was made elsewhere

11. How completely have your collaborating clinics/hospitals recorded data on the diagnosis of Peripheral Arterial Disease– this may or may not differ from what you yourselves are recording?

☐ These diagnoses have been collected systematically for all/most patients from a given point in time

☐ These diagnoses have been collected systematically for some patients from a given point in time

☐ These diagnoses have been collected but not systematically and/or recording is erratic

☐ We know that at least some clinics collect these diagnoses but we do have details of when they started and/or whether it is collected and recorded systematically.

12. What information are your collaborating clinics currently collecting relating to Peripheral Arterial Disease?

☐ ☐ ☐ Diagnosis Date

☐ ☐ ☐ Clinical Findings

☐ ☐ ☐ ICD-9/ICD-10 Codes

☐ ☐ ☐ Imaging results

☐ ☐ ☐ Ankle Brachial Index

☐ ☐ ☐ We do not know what they are collecting (would it be possible to find out?)

STROKE

(Including haemorrhage and infarction but EXCLUDING sub-arachnoid haemorrhage)

SECTION 11

1. Do you collect ANY data on Stroke?

☐ Yes (Go to **Question 3**)

☐ No

2. Do your collaborating clinics collect ANY data on Stroke?

☐ Yes (Go to **Question 11**)

☐ No (Go to **Section 12**)

☐ Don't Know- Would it be possible for you to find out? (Go to **Section 12**)

3. When did you start collecting data on Stroke?

4. Approximately how many Events have been recorded (if known)?

5. Approximately how many patient years of follow up have been recorded (if known)?

6. Has your Cohort determined a Standardized Case Definition(s) for Stroke?

☐ Yes

☐ No (if no go to **Question 9**)

7. Which of the following form part of your Case Definition?

☐ CT or MRI findings compatible with stroke diagnosis

☐ Clinical findings including a localizing neurological deficit

☐ Diagnostic codes (e.g. ICD-9/ICD-10)

☐ Other – please specify

8. Is there any central adjudication of these events through an end point review committee?

☐ Yes

☐ No

9. How comprehensively are you recording Stroke diagnoses?

☐ These diagnoses have been collected systematically for all/most patients from a given point in time

☐ These diagnoses have been collected systematically for some patients from a given point in time

☐ These diagnoses have been collected but not systematically and/or recording is erratic

10. What data are you recording relating to these events for each patient?

Yes	No	
☐	☐	Diagnosis Date
☐	☐	Clinical Findings
☐	☐	ICD-9/ICD-10 Codes
☐	☐	Imaging results
☐	☐	Case definition based splitting of events into confirmed and probable categories
☐	☐	Whether the diagnosis was made elsewhere

11. How completely have your collaborating clinics/hospitals recorded data on the diagnosis of Stroke – this may or may not differ from what you yourselves are recording?

☐ These diagnoses have been collected systematically for all/most patients from a given point in time

☐ These diagnoses have been collected systematically for some patients from a given point in time

☐ These diagnoses have been collected but not systematically and/or recording is erratic

☐ We know that at least some clinics collect these diagnoses but we do have details of when they started and/or whether it is collected and recorded systematically.

12. What information are your collaborating clinics currently collecting relating to Stroke?

All Clinics	Some Clinics	No Clinics	
☐	☐	☐	Diagnosis Date
☐	☐	☐	Clinical Findings
☐	☐	☐	ICD-9/ICD-10 Codes
☐	☐	☐	Imaging results
☐	☐	☐	Stroke type
☐	☐	☐	We do not know what they are collecting (would it be possible to find out?)

DEEP VEIN THROMBOSIS

SECTION 12

1. Do you collect ANY data on Deep Vein Thrombosis?

☐ Yes (Go to **Question 3**)

☐ No

2. Do your collaborating clinics collect ANY data on Deep Vein Thrombosis?

☐ Yes (Go to **Question 11**)

☐ No (Go to **Section 13)**

☐ Don't Know- Would it be possible to find out? (Go to **Section 13**)

3. When did you start collecting data on Deep Vein Thrombosis?

4. Approximately how many Events have been recorded (if known)?

5. Approximately how many patient years of follow up have been recorded (if known)?

6. Has your Cohort determined a Standardized Case Definition for Deep Vein Thrombosis?

☐ Yes

☐ No (if no go to **Question 9**)

7. Which of the following form part of your Case Definition?

☐ Diagnosis of DVT via imagining techniques e.g. contrast venography, helical computer tomography, MRI, ultrasonography

☐ Elevated D-dimer test

☐ Abnormal plethysmography

☐ Wells Clinical Prediction Rule for DVT

☐ No alternative diagnosis with greater likelihood than DVT

☐ Diagnostic Codes (e.g. ICD-9/ICD-10)

☐ Other – please specify

8. Is there any central adjudication of these events through an end point review committee?

☐ Yes

☐ No

9. How comprehensively are you recording Deep Vein Thrombosis diagnoses?

☐ These diagnoses have been collected systematically for all/most patients from a given point in time

☐ These diagnoses have been collected systematically for some patients from a given point in time

☐ These diagnoses have been collected but not systematically and/or recording is erratic

10. What data are you recording relating to these events for each patient?

Yes	No	
☐	☐	Diagnosis Date
☐	☐	Clinical Findings
☐	☐	ICD-9/ICD-10 Codes
☐	☐	Imaging results
☐	☐	Case definition based splitting of events into confirmed and probable categories
☐	☐	Whether the diagnosis was made elsewhere

11. How completely have your collaborating clinics/hospitals recorded data on the diagnosis of DVT– this may or may not differ from what you yourselves are recording?

☐ These diagnoses have been collected systematically for all/most patients from a given point in time

☐ These diagnoses have been collected systematically for some patients from a given point in time

☐ These diagnoses have been collected but not systematically and/or recording is erratic

☐ We know that at least some clinics collect these diagnoses but we do have details of when they started and/or whether it is collected and recorded systematically.

12. What information are your collaborating clinics currently collecting relating DVT?

All Clinics	Some Clinics	No Clinics	
☐	☐	☐	Diagnosis Date
☐	☐	☐	Clinical Findings
☐	☐	☐	ICD-9/ICD-10 Codes
☐	☐	☐	Imaging results
☐	☐	☐	We do not know what they are collecting (would it be possible to find out?)

ACUTE PANCREATITIS

SECTION 13

1. Do you collect ANY data on Acute Pancreatitis?

☐ Yes (Go to **Question 3**)

☐ No

2. Do your collaborating clinics collect ANY data on Acute Pancreatitis?

☐ Yes (Go to **Question 11**)

☐ No (Go to **END)**

☐ Don't Know (Go to **END**)

3. When did you start collecting data on Acute Pancreatitis?

4. Approximately how many Events have been recorded (if known)?

5. Approximately how many patient years of follow up have been recorded (if known)?

6. Has your Cohort determined a Standardized Case Definition for Acute Pancreatitis?

☐ Yes

☐ No (if no go to **Question 9**)

7. Which of the following form part of your Case Definition?

☐ Upper abdominal pain

☐ Elevated Lipase (at how many times normal levels?)

☐ Elevated Amylase (at how many times normal levels?)

☐ Ultrasound, CT or MRI suggesting Acute Pancreatitis

☐ Endoscopic ultrasound or Magnetic Resonance Cholangiopancreatography suggesting Acute Pancreatitis

☐ ICD-9/ICD-10 codes

☐ Other – please specify

8. Is there any central adjudication of these events through an end point review committee?

☐ Yes

☐ No

9. How comprehensively are you recording Acute Pancreatitis diagnoses?

☐ These diagnoses have been collected systematically for all/most patients from a given point in time

☐ These diagnoses have been collected systematically for some patients from a given point in time

☐ These diagnoses have been collected but not systematically and/or recording is erratic

10. What data are you recording relating to these events for each patient?

Yes	No	
☐	☐	Diagnosis Date
☐	☐	Clinical Findings
☐	☐	Enzyme Levels
☐	☐	Imaging results
☐	☐	ICD-9/ICD-10 Codes
☐	☐	Case definition based splitting of events into confirmed and probable categories

11. How completely have your collaborating clinics/hospitals recorded data on the diagnosis of Acute Pancreatitis– this may or may not differ from what you yourselves are recording?

☐ These diagnoses have been collected systematically for all/most patients from a given point in time

☐ These diagnoses have been collected systematically for some patients from a given point in time

☐ These diagnoses have been collected but not systematically and/or recording is erratic

☐ We know that at least some clinics collect these diagnoses but we do have details of when they started and/or whether it is collected and recorded systematically.

12. What information are your collaborating clinics currently collecting relating Acute Pancreatitis?

☐ ☐ ☐ Diagnosis Date

☐ ☐ ☐ Clinical Findings

☐ ☐ ☐ Enzyme Levels

☐ ☐ ☐ Imaging results

☐ ☐ ☐ ICD-9/ICD-10 Codes

☐ ☐ ☐ We do not know what they are collecting (would it be possible to find out?)

END

This is the end of the survey, thank you so much for taking the time to complete it. Please would you return it to me as an attachment and include any additional information which you think I might find useful.

Appendix C: Data Collection Form (HICDEP Format)

tblSNAE (All Serious non-AIDS Events)*

PATIENT	identifier for seroconverter	(string) comprises 3-letter cohort identifier followed by seroconverter *e.g.* UKR01001 *for UK Register patient 01001*unique identifier
BIRTH_D	Birth date	yyyy-mm-dd
SNAE_ID	Code to identify event	AMI Acute Myocardial Infarction DIA Diabetes Mellitus ESRD End-Stage Renal Disease ESLD End-Stage Liver Disease FRA Fracture NADM Non-AIDS defining Malignancy PAN Pancreatitis STR Stroke
NADM_ID	Code to identify nADM	ALL=Leukemia: Acute lymphoid AML=Leukemia: Acute myeloid ANAL=Anal cancer BLAD=Bladder cancer BRCA=Breast cancer CERV=Cervical dysplasia/carcinoma in situ CLL=Leukemia: Chronic lymphoid CML=Leukemia: Chronic myeloid COLO=Colon cancer COTC=Connective tissue cancer HDL=Hodgkin lymphoma KIDN=Kidney cancer LEUK=Leukemia: unspecified LIPC=Lip cancer LIVR=Liver cancer (HCC)

		LUNG=Lung cancer MALM=Malignant melanoma MEAC=Metastasis: of adenocarcinoma MEOC=Metastasis: of other cancertype MESC=Metastasis: of squamuos cell carcinoma META=Metastasis: unspecified MULM=Multiple myeloma PENC=Penile cancer PROS=Prostate cancer RECT=Rectum cancer STOM=Stomach cancer TESE=Testicular seminoma UTER=Uterus cancer
SNAE_D	date of event	yyyy-mm-dd

* this is based on the system of coding events previously developed and used by EuroSIDA

References

1. Sharp, P.M. and B.H. Hahn, *Origins of HIV and the AIDS pandemic.* Cold Spring Harb Perspect Med, 2011. **1**(1): p. a006841.
2. Lackner, A.A., M.M. Lederman, and B. Rodriguez, *HIV pathogenesis: the host.* Cold Spring Harb Perspect Med, 2012. **2**(9): p. a007005.
3. CDC. *Appendix A: AIDS-Defining Conditions* 2008; Available from: https://www.cdc.gov/mmwr/preview/mmwrhtml/rr5710a2.htm.
4. UNAIDS. *Fact Sheet – World AIDS Day 2017*. 2017 18/12/2017]; Available from: http://www.unaids.org/sites/default/files/media_asset/UNAIDS_FactSheet_en.pdf.
5. *Time from HIV-1 seroconversion to AIDS and death before widespread use of highly-active antiretroviral therapy: a collaborative re-analysis. Collaborative Group on AIDS Incubation and HIV Survival including the CASCADE EU Concerted Action. Concerted Action on SeroConversion to AIDS and Death in Europe.* Lancet, 2000. **355**(9210): p. 1131-7.
6. Mocroft, A., et al., *Survival after diagnosis of AIDS: a prospective observational study of 2625 patients. Royal Free/Chelsea and Westminster Hospitals Collaborative Group.* BMJ, 1997. **314**(7078): p. 409-13.
7. Sabin, C.A., *Do people with HIV infection have a normal life expectancy in the era of combination antiretroviral therapy?* BMC Med, 2013. **11**: p. 251.
8. May, M.T. and S.M. Ingle, *Life expectancy of HIV-positive adults: a review.* Sex Health, 2011. **8**(4): p. 526-33.
9. Antiretroviral Therapy Cohort Collaboration, *Survival of HIV-positive patients starting antiretroviral therapy between 1996 and 2013: a collaborative analysis of cohort studies.* Lancet HIV, 2017. **4**(8): p. e349-e356.
10. Patterson, S., et al., *Life expectancy of HIV-positive individuals on combination antiretroviral therapy in Canada.* BMC Infect Dis, 2015. **15**: p. 274.
11. Samji, H., et al., *Closing the gap: increases in life expectancy among treated HIV-positive individuals in the United States and Canada.* PLoS One, 2013. **8**(12): p. e81355.
12. Mocroft, A., et al., *Changes in the cause of death among HIV positive subjects across Europe: results from the EuroSIDA study.* AIDS, 2002. **16**(12): p. 1663-71.
13. Bonnet, F., et al., *Causes of death among HIV-infected patients in the era of highly active antiretroviral therapy, Bordeaux, France, 1998-1999.* HIV Med, 2002. **3**(3): p. 195-9.
14. Smit, C., et al., *Effective therapy has altered the spectrum of cause-specific mortality following HIV seroconversion.* AIDS, 2006. **20**(5): p. 741-9.
15. Palella, F.J., et al., *Mortality in the highly active antiretroviral therapy era: changing causes of death and disease in the HIV outpatient study.* J Acquir Immune Defic Syndr, 2006. **43**(1): p. 27-34.
16. Crum, N.F., et al., *Comparisons of causes of death and mortality rates among HIV-infected persons: analysis of the pre-, early, and late HAART (highly active antiretroviral therapy) eras.* J Acquir Immune Defic Syndr, 2006. **41**(2): p. 194-200.
17. Lau, B., S.J. Gange, and R.D. Moore, *Risk of non-AIDS-related mortality may exceed risk of AIDS-related mortality among individuals enrolling into care with CD4+ counts greater than 200 cells/mm3.* J Acquir Immune Defic Syndr, 2007. **44**(2): p. 179-87.
18. Belloso, W.H., et al., *Analysis of serious non-AIDS events among HIV-infected adults at Latin American sites.* HIV Med, 2010. **11**(9): p. 554-64.

19. Lifson, A.R., et al., *Development of diagnostic criteria for serious non-AIDS events in HIV clinical trials.* HIV Clin Trials, 2010. **11**(4): p. 205-19.
20. Neuhaus, J., et al., *Risk of all-cause mortality associated with nonfatal AIDS and serious non-AIDS events among adults infected with HIV.* AIDS, 2010. **24**(5): p. 697-706.
21. *Causes of death in HIV-1-infected patients treated with antiretroviral therapy, 1996-2006: collaborative analysis of 13 HIV cohort studies.* Clin Infect Dis, 2010. **50**(10): p. 1387-96.
22. Louie, J.K., et al., *Trends in causes of death among persons with acquired immunodeficiency syndrome in the era of highly active antiretroviral therapy, San Francisco, 1994-1998.* J Infect Dis, 2002. **186**(7): p. 1023-7.
23. Sackoff, J.E., et al., *Causes of death among persons with AIDS in the era of highly active antiretroviral therapy: New York City.* Ann Intern Med, 2006. **145**(6): p. 397-406.
24. Weber, R., et al., *Decreasing mortality and changing patterns of causes of death in the Swiss HIV Cohort Study.* HIV Med, 2013. **14**(4): p. 195-207.
25. Anis, A.H., et al., *Quality of life of patients with advanced HIV/AIDS: measuring the impact of both AIDS-defining events and non-AIDS serious adverse events.* J Acquir Immune Defic Syndr, 2009. **51**(5): p. 631-9.
26. Grulich, A.E., et al., *Incidence of cancers in people with HIV/AIDS compared with immunosuppressed transplant recipients: a meta-analysis.* Lancet, 2007. **370**(9581): p. 59-67.
27. Thio, C.L., et al., *HIV-1, hepatitis B virus, and risk of liver-related mortality in the Multicenter Cohort Study (MACS).* Lancet, 2002. **360**(9349): p. 1921-6.
28. Choi, A.I., et al., *Racial differences in end-stage renal disease rates in HIV infection versus diabetes.* J Am Soc Nephrol, 2007. **18**(11): p. 2968-74.
29. Darby, S.C., et al., *Mortality from liver cancer and liver disease in haemophilic men and boys in UK given blood products contaminated with hepatitis C. UK Haemophilia Centre Directors' Organisation.* Lancet, 1997. **350**(9089): p. 1425-31.
30. Triant, V.A., et al., *Increased acute myocardial infarction rates and cardiovascular risk factors among patients with human immunodeficiency virus disease.* J Clin Endocrinol Metab, 2007. **92**(7): p. 2506-12.
31. Hymes, K.B., et al., *Kaposi's sarcoma in homosexual men-a report of eight cases.* Lancet, 1981. **2**(8247): p. 598-600.
32. Masur, H., et al., *An outbreak of community-acquired Pneumocystis carinii pneumonia: initial manifestation of cellular immune dysfunction.* N Engl J Med, 1981. **305**(24): p. 1431-8.
33. Haverkos, H.W. and J.W. Curran, *The current outbreak of Kaposi's sarcoma and opportunistic infections.* CA Cancer J Clin, 1982. **32**(6): p. 330-9.
34. Barre-Sinoussi, F., et al., *Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS).* Science, 1983. **220**(4599): p. 868-71.
35. Gallo, R.C., et al., *Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS).* Science, 1983. **220**(4599): p. 865-7.
36. Ratner, L., R.C. Gallo, and F. Wong-Staal, *HTLV-III, LAV, ARV are variants of same AIDS virus.* Nature, 1985. **313**(6004): p. 636-7.
37. AVERT. *HIV Strains and Types* 2016; Available from: http://www.avert.org/professionals/hiv-science/types-strains.
38. Junqueira, D.M. and S.E.d.M. Almeida, *HIV-1 subtype B: Traces of a pandemic.* Virology, 2016. **495**: p. 173-184.

39. Merck, *Simplified HIV life cycle* 2017; Available from: http://www.merckmanuals.com/professional/infectious-diseases/human-immunodeficiency-virus-hiv/human-immunodeficiency-virus-hiv-infection.
40. Shaw, G.M. and E. Hunter, *HIV transmission.* Cold Spring Harb Perspect Med, 2012. **2**(11).
41. AIDSinfo. *The HIV Life Cycle* 2017; Available from: http://med.unr.edu/Documents/unsom/statewide/aidsetc/HIVLifeCycle_FS_en.pdf .
42. Simon, V., D. Ho, and Q. Abdool Karim, *HIV/AIDS epidemiology, pathogenesis, prevention, and treatment.* Lancet, 2006. **368**(9534): p. 489-504.
43. Swanstrom, R. and J. Coffin, *HIV-1 pathogenesis: the virus.* Cold Spring Harb Perspect Med, 2012. **2**(12): p. a007443.
44. Samson, M., et al., *Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene.* Nature, 1996. **382**(6593): p. 722-5.
45. AIDSMAP *Factsheet: Primary Infection* Available from: http://www.aidsmap.com/Primary-infection/page/1044761/.
46. AIDS.GOV *Stages of HIV Infection* Available from: https://www.aids.gov/hiv-aids-basics/just-diagnosed-with-hiv-aids/hiv-in-your-body/stages-of-hiv/.
47. Brenchley, J.M., et al., *CD4+ T cell depletion during all stages of HIV disease occurs predominantly in the gastrointestinal tract.* J Exp Med, 2004. **200**(6): p. 749-59.
48. Li, Q., et al., *Peak SIV replication in resting memory CD4+ T cells depletes gut lamina propria CD4+ T cells.* Nature, 2005. **434**(7037): p. 1148-52.
49. Rodgers, V.D., R. Fassett, and M.F. Kagnoff, *Abnormalities in intestinal mucosal T cells in homosexual populations including those with the lymphadenopathy syndrome and acquired immunodeficiency syndrome.* Gastroenterology, 1986. **90**(3): p. 552-8.
50. Lyles, R.H., et al., *Natural history of human immunodeficiency virus type 1 viremia after seroconversion and proximal to AIDS in a large cohort of homosexual men. Multicenter AIDS Cohort Study.* J Infect Dis, 2000. **181**(3): p. 872-80.
51. Hubert, J.B., et al., *Natural history of serum HIV-1 RNA levels in 330 patients with a known date of infection. The SEROCO Study Group.* AIDS, 2000. **14**(2): p. 123-31.
52. Soriano, V., et al., *The decline in CD4+ T lymphocytes as a function of the duration of HIV infection, age at seroconversion, and viral load.* J Infect, 1998. **36**(3): p. 307-11.
53. Rodriguez, B., et al., *Predictive value of plasma HIV RNA level on rate of CD4 T-cell decline in untreated HIV infection.* JAMA, 2006. **296**(12): p. 1498-506.
54. Platt, L., et al., *Prevalence and burden of HCV co-infection in people living with HIV: a global systematic review and meta-analysis.* Lancet Infect Dis, 2016. **16**(7): p. 797-808.
55. Kourtis, A.P., et al., *HIV-HBV coinfection--a global challenge.* N Engl J Med, 2012. **366**(19): p. 1749-52.
56. Chen, J.Y., E.R. Feeney, and R.T. Chung, *HCV and HIV co-infection: mechanisms and management.* Nat Rev Gastroenterol Hepatol, 2014. **11**(6): p. 362-71.
57. Weber, R., et al., *Liver-related deaths in persons infected with the human immunodeficiency virus: the D:A:D study.* Arch Intern Med, 2006. **166**(15): p. 1632-41.
58. Cooke, G.S., et al., *Viral hepatitis and the Global Burden of Disease: a need to regroup.* J Viral Hepat, 2013. **20**(9): p. 600-1.

59. Thrift, A.P., H.B. El-Serag, and F. Kanwal, *Global epidemiology and burden of HCV infection and HCV-related disease.* Nat Rev Gastroenterol Hepatol, 2017. **14**(2): p. 122-132.
60. Peters, L., et al., *HIV and hepatitis C co-infection in Europe, Israel and Argentina: a EuroSIDA perspective.* BMC Infect Dis, 2014. **14 Suppl 6**: p. S13.
61. Rauch, A., et al., *Unsafe sex and increased incidence of hepatitis C virus infection among HIV-infected men who have sex with men: the Swiss HIV Cohort Study.* Clin Infect Dis, 2005. **41**(3): p. 395-402.
62. Mohsen, A.H., S. Murad, and P.J. Easterbrook, *Prevalence of hepatitis C in an ethnically diverse HIV-1-infected cohort in south London.* HIV Med, 2005. **6**(3): p. 206-15.
63. Boesecke, C., et al., *Hepatitis C seroconversions in HIV infection across Europe: which regions and patient groups are affected?* Liver Int, 2015. **35**(11): p. 2384-91.
64. Pakianathan, M., et al., *Chemsex and new HIV diagnosis in gay, bisexual and other men who have sex with men attending sexual health clinics.* HIV Med, 2018.
65. Westbrook, R.H. and G. Dusheiko, *Natural history of hepatitis C.* J Hepatol, 2014. **61**(1 Suppl): p. S58-68.
66. Rotman, Y. and T.J. Liang, *Coinfection with hepatitis C virus and human immunodeficiency virus: virological, immunological, and clinical outcomes.* J Virol, 2009. **83**(15): p. 7366-74.
67. Thio, C.L., *Hepatitis B and human immunodeficiency virus coinfection.* Hepatology, 2009. **49**(5 Suppl): p. S138-45.
68. MacLachlan, J.H. and B.C. Cowie, *Hepatitis B virus epidemiology.* Cold Spring Harb Perspect Med, 2015. **5**(5): p. a021410.
69. 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.
70. 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.
71. Flynn, N.M., et al., *Placebo-controlled phase 3 trial of a recombinant glycoprotein 120 vaccine to prevent HIV-1 infection.* J Infect Dis, 2005. **191**(5): p. 654-65.
72. Priddy, F.H., et al., *Safety and immunogenicity of a replication-incompetent adenovirus type 5 HIV-1 clade B gag/pol/nef vaccine in healthy adults.* Clin Infect Dis, 2008. **46**(11): p. 1769-81.
73. Buchbinder, S.P., et al., *Efficacy assessment of a cell-mediated immunity HIV-1 vaccine (the Step Study): a double-blind, randomised, placebo-controlled, test-of-concept trial.* Lancet, 2008. **372**(9653): p. 1881-93.
74. Gray, G.E., et al., *Safety and efficacy of the HVTN 503/Phambili study of a clade-B-based HIV-1 vaccine in South Africa: a double-blind, randomised, placebo-controlled test-of-concept phase 2b study.* Lancet Infect Dis, 2011. **11**(7): p. 507-15.
75. *Imbokodo vaccine study* 2017 29/12/2017]; Available from: https://www.imbokodo.org.za/faq.
76. US Food and Drug Administration. *Antiretroviral drugs used in the treatment of HIV infection*. 2018; Available from: https://www.fda.gov/forpatients/illness/hivaids/treatment/ucm118915.htm.
77. European AIDS Clinical Society *Guidelines: Version 9.1 (English)*. 2018; Available from: http://www.eacsociety.org/files/2018_guidelines-9.1-english.pdf.
78. Arts, E.J. and D.J. Hazuda, *HIV-1 antiretroviral drug therapy.* Cold Spring Harb Perspect Med, 2012. **2**(4): p. a007161.

79. Brook, I., *Approval of zidovudine (AZT) for acquired immunodeficiency syndrome. A challenge to the medical and pharmaceutical communities.* JAMA, 1987. **258**(11): p. 1517.
80. *Advances in the management of HIV disease: focus on didanosine (ddI). A symposium. Houston, Texas, 18 September 1989. Proceedings.* Rev Infect Dis, 1990. **12 Suppl 5**: p. S499-576.
81. Lambert, J.S., et al., *2',3'-dideoxyinosine (ddI) in patients with the acquired immunodeficiency syndrome or AIDS-related complex. A phase I trial.* N Engl J Med, 1990. **322**(19): p. 1333-40.
82. *Zidovudine, didanosine, and zalcitabine in the treatment of HIV infection: meta-analyses of the randomised evidence. HIV Trialists' Collaborative Group.* Lancet, 1999. **353**(9169): p. 2014-25.
83. Cheeseman, S.H., et al., *Phase I/II evaluation of nevirapine alone and in combination with zidovudine for infection with human immunodeficiency virus.* J Acquir Immune Defic Syndr Hum Retrovirol, 1995. **8**(2): p. 141-51.
84. Kitchen, V.S., et al., *Safety and activity of saquinavir in HIV infection.* Lancet, 1995. **345**(8955): p. 952-5.
85. Mocroft, A., et al., *Changing patterns of mortality across Europe in patients infected with HIV-1. EuroSIDA Study Group.* Lancet, 1998. **352**(9142): p. 1725-30.
86. Palella, F.J., Jr., et al., *Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. HIV Outpatient Study Investigators.* N Engl J Med, 1998. **338**(13): p. 853-60.
87. Ene, L., D. Duiculescu, and S.M. Ruta, *How much do antiretroviral drugs penetrate into the central nervous system?* J Med Life, 2011. **4**(4): p. 432-9.
88. Llibre, J.M., et al., *PHASE III SWORD 1&2: SWITCH TO DTG+RPV MAINTAINS VIROLOGIC SUPPRESSION THROUGH 48 WKS*, in *CROI*. 2017: WA: Seattle
89. Blair, H.A., *Dolutegravir/Rilpivirine: A Review in HIV-1 Infection.* Drugs, 2018. **78**(16): p. 1741-1750.
90. British HIV Association, *Guidelines for the treatment of HIV-1-positive adults with antiretroviral therapy 2012 (Updated November 2013).* HIV Med, 2014. **15 Suppl 1**: p. 1-85.
91. Group, I.S.S., et al., *Initiation of Antiretroviral Therapy in Early Asymptomatic HIV Infection.* N Engl J Med, 2015. **373**(9): p. 795-807.
92. Opravil, M., et al., *Clinical efficacy of early initiation of HAART in patients with asymptomatic HIV infection and CD4 cell count > 350 x 10(6) /l.* AIDS, 2002. **16**(10): p. 1371-81.
93. When To Start Consortium, *Timing of initiation of antiretroviral therapy in AIDS-free HIV-1-infected patients: a collaborative analysis of 18 HIV cohort studies.* Lancet, 2009. **373**(9672): p. 1352-63.
94. The Cascade Collaboration, *Timing of HAART initiation and clinical outcomes in human immunodeficiency virus type 1 seroconverters.* Arch Intern Med, 2011. **171**(17): p. 1560-9.
95. European AIDS Clinical Society, *Guidelines Archive*. 2018; Available from: http://www.eacsociety.org/guidelines/guidelines-archive/archive.html.
96. The TEMPRANO ANRS Study Group, et al., *A Trial of Early Antiretrovirals and Isoniazid Preventive Therapy in Africa.* N Engl J Med, 2015. **373**(9): p. 808-22.
97. Lundgren, J.D., et al., *Inferior clinical outcome of the CD4+ cell count-guided antiretroviral treatment interruption strategy in the SMART study: role of CD4+ Cell counts and HIV RNA levels during follow-up.* J Infect Dis, 2008. **197**(8): p. 1145-55.
98. Cohen, M.S., et al., *Prevention of HIV-1 infection with early antiretroviral therapy.* N Engl J Med, 2011. **365**(6): p. 493-505.

99. The HIV-Causal Collaboration, et al., *The effect of combined antiretroviral therapy on the overall mortality of HIV-infected individuals.* AIDS, 2010. **24**(1): p. 123-37.
100. Kitahata, M.M., et al., *Effect of early versus deferred antiretroviral therapy for HIV on survival.* N Engl J Med, 2009. **360**(18): p. 1815-26.
101. Palella, F.J., Jr., et al., *Survival benefit of initiating antiretroviral therapy in HIV-infected persons in different CD4+ cell strata.* Ann Intern Med, 2003. **138**(8): p. 620-6.
102. World Health Organisation, *Consolidated Guidelines on The Use Of Antiretroviral Drugs For Treating And Preventing HIV Infection: Recommendations for a Public Health Approach: Second Edition 2016* 29/12/2017]; Available from: http://apps.who.int/iris/bitstream/10665/208825/1/9789241549684_eng.pdf?ua=1.
103. World Health Organisation, *Updated recommendations on first-line and second-line antiretroviral regimens and post-exposure prophylaxis and recommendations on early infant diagnosis of HIV.* 2018.
104. British HIV Association, *guidelines for the treatment of HIV-1-positive adults with antiretroviral therapy 2015 (2016 interim update)* 2017 29/12/2017]; Available from: http://www.bhiva.org/HIV-1-treatment-guidelines.aspx.
105. U.S. Department of Health and Human Services, *Guidelines for the Use of Antiretroviral Agents in Adults and Adolescents Living with HIV*; Available from: https://aidsinfo.nih.gov/guidelines/html/1/adult-and-adolescent-treatment-guidelines/0.
106. Panel on Antiretroviral Guidelines for Adults and Adolescents, *Guidelines for the Use of Antiretroviral Agents in Adults andAdolescents with HIV*. 2018 09/05/2019]; Available from: https://aidsinfo.nih.gov/contentfiles/lvguidelines/AdultandAdolescentGL.pdf.
107. UNAIDS, *Fact Sheet: Latest statistics on the status of the AIDS epidemic*. 2017; Available from: http://www.unaids.org/en/resources/fact-sheet.
108. May, M., et al., *Impact of late diagnosis and treatment on life expectancy in people with HIV-1: UK Collaborative HIV Cohort (UK CHIC) Study.* BMJ, 2011. **343**: p. d6016.
109. Masia, M., et al., *Risk, predictors, and mortality associated with non-AIDS events in newly diagnosed HIV-infected patients: role of antiretroviral therapy.* AIDS, 2013. **27**(2): p. 181-9.
110. Baker, J.V., et al., *CD4+ count and risk of non-AIDS diseases following initial treatment for HIV infection.* AIDS, 2008. **22**(7): p. 841-8.
111. Reekie, J., et al., *Fatal and nonfatal AIDS and non-AIDS events in HIV-1-positive individuals with high CD4 cell counts according to viral load strata.* AIDS, 2011. **25**(18): p. 2259-68.
112. Masia, M., et al., *Risk, predictors, and mortality associated with non-AIDS events in newly diagnosed HIV-infected patients: role of antiretroviral therapy.* AIDS, 2013. **27**(2): p. 181-9.
113. Lucero, C., et al., *Rate and predictors of non-AIDS events in a cohort of HIV-infected patients with a CD4 T cell count above 500 cells/mm(3).* AIDS Res Hum Retroviruses, 2013. **29**(8): p. 1161-7.
114. Bonnet, F., et al., *Trends and determinants of severe morbidity in HIV-infected patients: the ANRS CO3 Aquitaine Cohort, 2000-2004.* HIV Med, 2007. **8**(8): p. 547-54.
115. Reisler, R.B., et al., *Grade 4 events are as important as AIDS events in the era of HAART.* J Acquir Immune Defic Syndr, 2003. **34**(4): p. 379-86.

116. World Health Organisation, *Leading Causes of Death by Economy Income Group*; Available from: http://www.who.int/mediacentre/factsheets/fs310/en/index1.html.
117. Phillips, A.N., J. Neaton, and J.D. Lundgren, *The role of HIV in serious diseases other than AIDS.* AIDS, 2008. **22**(18): p. 2409-18.
118. Islam, F.M., et al., *Relative risk of cardiovascular disease among people living with HIV: a systematic review and meta-analysis.* HIV Med, 2012. **13**(8): p. 453-68.
119. Shiau, S., et al., *Incident fractures in HIV-infected individuals: a systematic review and meta-analysis.* AIDS, 2013. **27**(12): p. 1949-57.
120. Shiels, M.S., et al., *A meta-analysis of the incidence of non-AIDS cancers in HIV-infected individuals.* J Acquir Immune Defic Syndr, 2009. **52**(5): p. 611-22.
121. Drozd, D.R., et al., *Increased Risk of Myocardial Infarction in HIV-Infected Individuals in North America Compared With the General Population.* J Acquir Immune Defic Syndr, 2017. **75**(5): p. 568-576.
122. Rafiq, N., et al., *Predictors of chronic liver disease in individuals with human immunodeficiency virus infection.* Ann Hepatol, 2013. **13**(1): p. 60-4.
123. Rasmussen, L.D., et al., *Risk of diabetes mellitus in persons with and without HIV: a Danish nationwide population-based cohort study.* PLoS One, 2012. **7**(9): p. e44575.
124. Brown, T.T., et al., *Antiretroviral therapy and the prevalence and incidence of diabetes mellitus in the multicenter AIDS cohort study.* Arch Intern Med, 2005. **165**(10): p. 1179-84.
125. Rao, S.G., et al., *Factors Associated With Excess Myocardial Infarction Risk in HIV-Infected Adults: A Systematic Review and Meta-analysis.* J Acquir Immune Defic Syndr, 2019. **81**(2): p. 224-230.
126. Warriner, M.D., et al., *Fracture among Older and Younger HIV+ Medicare Beneficiaries* in *Confrence on Retroviruses and Opportunistic Infections* 2013: Atlanta, Georgia
127. Arnsten, J.H., et al., *Decreased bone mineral density and increased fracture risk in aging men with or at risk for HIV infection.* AIDS, 2007. **21**(5): p. 617-23.
128. Sharma, A., et al., *Fracture Incidence is Increased in Aging HIV-Infected Women* in *CROI 2015* 2015: Seattle,USA.
129. Walker-Harris, V., et al., *Incident bone fractures in men with, or at risk for, HIV-infection in the Multicenter AIDS Cohort Study (MACS), 1996-2011*, in *XIX International AIDS Conference*. 2012: Washington, DC. .
130. Guerri-Fernandez, R., et al., *HIV infection is strongly associated with hip fracture risk, independently of age, gender, and comorbidities: a population-based cohort study.* J Bone Miner Res, 2013. **28**(6): p. 1259-63.
131. Hansen, A.B., et al., *Incidence of low and high-energy fractures in persons with and without HIV infection: a Danish population-based cohort study.* Aids, 2012. **26**(3): p. 285-93.
132. Womack, J.A., et al., *Increased risk of fragility fractures among HIV infected compared to uninfected male veterans.* PLoS One, 2011. **6**(2): p. e17217.
133. Brar, I., et al., *A comparison of factors associated with prevalent diabetes mellitus among HIV-Infected antiretroviral-naive individuals versus individuals in the National Health and Nutritional Examination Survey cohort.* J Acquir Immune Defic Syndr, 2007. **45**(1): p. 66-71.
134. Butt, A.A., et al., *HIV infection and the risk of diabetes mellitus.* AIDS, 2009. **23**(10): p. 1227-34.
135. Smith, C., et al., *Factors associated with specific causes of death amongst HIV-positive individuals in the D:A:D Study.* AIDS, 2010. **24**(10): p. 1537-48.

136. Palella, F.J., et al., *Mortality in the highly active antiretroviral therapy era: changing causes of death and disease in the HIV outpatient study.* J Acquir Immune Defic Syndr, 2006. **43**(1): p. 27-34.
137. Grulich, A.E., et al., *Rates of non-AIDS-defining cancers in people with HIV infection before and after AIDS diagnosis.* AIDS, 2002. **16**(8): p. 1155-61.
138. Bedimo, R.J., et al., *Incidence of non-AIDS-defining malignancies in HIV-infected versus noninfected patients in the HAART era: impact of immunosuppression.* J Acquir Immune Defic Syndr, 2009. **52**(2): p. 203-8.
139. Powles, T., et al., *Highly active antiretroviral therapy and the incidence of non-AIDS-defining cancers in people with HIV infection.* J Clin Oncol, 2009. **27**(6): p. 884-90.
140. Alter, M.J., *Epidemiology of viral hepatitis and HIV co-infection.* J Hepatol, 2006. **44**(1 Suppl): p. S6-9.
141. Goulet, J.L., et al., *Aging and infectious diseases: do patterns of comorbidity vary by HIV status, age, and HIV severity?* Clin Infect Dis, 2007. **45**(12): p. 1593-601.
142. Deayton, J.R., et al., *Importance of cytomegalovirus viraemia in risk of disease progression and death in HIV-infected patients receiving highly active antiretroviral therapy.* Lancet, 2004. **363**(9427): p. 2116-21.
143. Lichtner, M., et al., *Cytomegalovirus coinfection is associated with an increased risk of severe non-AIDS-defining events in a large cohort of HIV-infected patients.* J Infect Dis, 2015. **211**(2): p. 178-86.
144. Wang, H., et al., *Cytomegalovirus Infection and Relative Risk of Cardiovascular Disease (Ischemic Heart Disease, Stroke, and Cardiovascular Death): A Meta-Analysis of Prospective Studies Up to 2016.* J Am Heart Assoc, 2017. **6**(7).
145. Herndier, B.G., et al., *High prevalence of Epstein-Barr virus in the Reed-Sternberg cells of HIV-associated Hodgkin's disease.* Am J Pathol, 1993. **142**(4): p. 1073-9.
146. Mehta, S.H., et al., *The effect of HAART and HCV infection on the development of hyperglycemia among HIV-infected persons.* J Acquir Immune Defic Syndr, 2003. **33**(5): p. 577-84.
147. Carter, M., *Focus on hepatitis. HIV/HCV coinfection and CVD risk.* IAPAC Mon, 2007. **13**(1): p. 27.
148. Loko, M.A., et al., *The French national prospective cohort of patients co-infected with HIV and HCV (ANRS CO13 HEPAVIH): early findings, 2006-2010.* BMC Infect Dis, 2010. **10**: p. 303.
149. Hernando, V., et al., *All-cause and liver-related mortality in HIV positive subjects compared to the general population: differences by HCV co-infection.* J Hepatol, 2012. **57**(4): p. 743-51.
150. van der Helm, J., et al., *Effect of HCV infection on cause-specific mortality after HIV seroconversion, before and after 1997.* Gastroenterology, 2013. **144**(4): p. 751-760 e2.
151. Carr, A., et al., *Diagnosis, prediction, and natural course of HIV-1 protease-inhibitor-associated lipodystrophy, hyperlipidaemia, and diabetes mellitus: a cohort study.* Lancet, 1999. **353**(9170): p. 2093-9.
152. Friis-Moller, N., et al., *Cardiovascular disease risk factors in HIV patients--association with antiretroviral therapy. Results from the DAD study.* AIDS, 2003. **17**(8): p. 1179-93.
153. Hughes, C.A., et al., *Risk factors for new-onset diabetes mellitus in patients receiving protease inhibitor therapy.* Can J Infect Dis Med Microbiol, 2005. **16**(4): p. 230-2.
154. Bedimo, R., et al., *Osteoporotic fracture risk associated with cumulative exposure to tenofovir and other antiretroviral agents.* AIDS, 2012. **26**(7): p. 825-31.

155. Worm, S.W., et al., *Risk of myocardial infarction in patients with HIV infection exposed to specific individual antiretroviral drugs from the 3 major drug classes: the data collection on adverse events of anti-HIV drugs (D:A:D) study.* J Infect Dis, 2010. **201**(3): p. 318-30.
156. Durand, M., et al., *Association between HIV infection, antiretroviral therapy, and risk of acute myocardial infarction: a cohort and nested case-control study using Quebec's public health insurance database.* J Acquir Immune Defic Syndr, 2011. **57**(3): p. 245-53.
157. Strategies for Management of Anti-Retroviral Therapy and D.A.D. Study Groups, *Use of nucleoside reverse transcriptase inhibitors and risk of myocardial infarction in HIV-infected patients.* AIDS, 2008. **22**(14): p. F17-24.
158. De Wit, S., et al., *Incidence and risk factors for new-onset diabetes in HIV-infected patients: the Data Collection on Adverse Events of Anti-HIV Drugs (D:A:D) study.* Diabetes Care, 2008. **31**(6): p. 1224-9.
159. Lobo, F.S., et al., *Addressing the issue of channeling bias in observational studies with propensity scores analysis.* Res Social Adm Pharm, 2006. **2**(1): p. 143-51.
160. Mark, S.D. and J.M. Robins, *Estimating the causal effect of smoking cessation in the presence of confounding factors using a rank preserving structural failure time model.* Stat Med, 1993. **12**(17): p. 1605-28.
161. El-Sadr, W.M., et al., *CD4+ count-guided interruption of antiretroviral treatment.* N Engl J Med, 2006. **355**(22): p. 2283-96.
162. Strategies for Management of Antiretroviral Therapy Study Group, *CD4+ count-guided interruption of antiretroviral treatment.* N Engl J Med, 2006. **355**(22): p. 2283-96.
163. Silverberg, M.J., et al., *Risk of cancers during interrupted antiretroviral therapy in the SMART study.* AIDS, 2007. **21**(14): p. 1957-63.
164. Carr, A., et al., *Prevalence of and risk factors for low bone mineral density in untreated HIV infection: a substudy of the INSIGHT Strategic Timing of AntiRetroviral Treatment (START) trial.* HIV Med, 2015. **16 Suppl 1**: p. 137-46.
165. Mocroft, A., et al., *Serious fatal and nonfatal non-AIDS-defining illnesses in Europe.* J Acquir Immune Defic Syndr, 2010. **55**(2): p. 262-70.
166. Grabar, S., et al., *Causes of the first AIDS-defining illness and subsequent survival before and after the advent of combined antiretroviral therapy.* HIV Med, 2008. **9**(4): p. 246-56.
167. Hsu, D.C., I. Sereti, and J. Ananworanich, *Serious Non-AIDS events: Immunopathogenesis and interventional strategies.* AIDS Res Ther, 2013. **10**(1): p. 29.
168. Neuhaus, J., et al., *Markers of inflammation, coagulation, and renal function are elevated in adults with HIV infection.* J Infect Dis, 2010. **201**(12): p. 1788-95.
169. Kuller, L.H., et al., *Inflammatory and coagulation biomarkers and mortality in patients with HIV infection.* PLoS Med, 2008. **5**(10): p. e203.
170. Jong, E., et al., *The effect of initiating combined antiretroviral therapy on endothelial cell activation and coagulation markers in South African HIV-infected individuals.* Thromb Haemost, 2010. **104**(6): p. 1228-34.
171. Somsouk, M., et al., *Gut epithelial barrier and systemic inflammation during chronic HIV infection.* AIDS, 2015. **29**(1): p. 43-51.
172. Mehandru, S., et al., *Mechanisms of gastrointestinal CD4+ T-cell depletion during acute and early human immunodeficiency virus type 1 infection.* J Virol, 2007. **81**(2): p. 599-612.

173. Guiguet, M., et al., *Effect of immunodeficiency, HIV viral load, and antiretroviral therapy on the risk of individual malignancies (FHDH-ANRS CO4): a prospective cohort study.* Lancet Oncol, 2009. **10**(12): p. 1152-9.
174. Sabin, C.A., et al., *Associations between immune depression and cardiovascular events in HIV infection.* Aids, 2013. **27**(17): p. 2735-48.
175. Lichtenstein, K.A., et al., *Low CD4+ T cell count is a risk factor for cardiovascular disease events in the HIV outpatient study.* Clin Infect Dis, 2010. **51**(4): p. 435-47.
176. Achhra, A.C., et al., *Immunodeficiency and the risk of serious clinical endpoints in a well studied cohort of treated HIV-infected patients.* AIDS, 2010. **24**(12): p. 1877-86.
177. Ibrahim, F., et al., *Immunodeficiency and renal impairment are risk factors for HIV-associated acute renal failure.* AIDS, 2010. **24**(14): p. 2239-44.
178. Yong, M.K., et al., *Low CD4 count is associated with an increased risk of fragility fracture in HIV-infected patients.* J Acquir Immune Defic Syndr, 2011. **57**(3): p. 205-10.
179. Opportunistic Infections Project Team of the Collaboration of Observational HIV Epidemiological Research Europe, *CD4 cell count and the risk of AIDS or death in HIV-Infected adults on combination antiretroviral therapy with a suppressed viral load: a longitudinal cohort study from COHERE.* PLoS Med, 2012. **9**(3): p. e1001194.
180. *World Bank: High Income Countries*; Available from: https://data.worldbank.org/income-level/high-income.
181. *The International AIDS Society: Abstract Archive*; Available from: http://www.abstract-archive.org/.
182. Young, B., et al., *Increased rates of bone fracture among HIV-infected persons in the HIV Outpatient Study (HOPS) compared with the US general population, 2000-2006.* Clin Infect Dis, 2011. **52**(8): p. 1061-8.
183. Battalora, L., K. Buchaxz, and C. Armon, *Low bone mineral density is associated with increased risk of incident fracture in HIV-infected adults* in *14th European AIDS Conference* 2013: Brussels
184. Womack, J.A., et al., *Physiologic frailty and fragility fracture in HIV-infected male veterans.* Clin Infect Dis, 2013. **56**(10): p. 1498-504.
185. Hasse, B., et al., *Morbidity and aging in HIV-infected persons: the Swiss HIV cohort study.* Clin Infect Dis, 2011. **53**(11): p. 1130-9.
186. Yin, M.T., et al., *Fracture incidence in HIV-infected women: results from the Women's Interagency HIV Study.* Aids, 2010. **24**(17): p. 2679-86.
187. Yin, M.T., et al., *Fractures after antiretroviral initiation.* AIDS, 2012. **26**(17): p. 2175-84.
188. Castronuovo, D., et al., *Bone disease in the setting of HIV infection: update and review of the literature.* Eur Rev Med Pharmacol Sci, 2013. **17**(18): p. 2413-9.
189. Kanis, J.A., et al., *Smoking and fracture risk: a meta-analysis.* Osteoporos Int, 2005. **16**(2): p. 155-62.
190. Kanis, J.A., et al., *A meta-analysis of previous fracture and subsequent fracture risk.* Bone, 2004. **35**(2): p. 375-82.
191. Kanis, J.A., et al., *A meta-analysis of prior corticosteroid use and fracture risk.* J Bone Miner Res, 2004. **19**(6): p. 893-9.
192. Kanis, J.A., et al., *A family history of fracture and fracture risk: a meta-analysis.* Bone, 2004. **35**(5): p. 1029-37.
193. Kanis, J.A., et al., *Alcohol intake as a risk factor for fracture.* Osteoporos Int, 2005. **16**(7): p. 737-42.
194. Maalouf, N.M., et al., *Hepatitis C co-infection and severity of liver disease as risk factors for osteoporotic fractures among HIV-infected patients.* J Bone Miner Res, 2013. **28**(12): p. 2577-83.

195. Gonciulea, A., et al., *An increased rate of fracture occurs a decade earlier in HIV plus compared with HIV-men.* Aids, 2017. **31**(10): p. 1435-1443.
196. Sharma, A., et al., *Increased Fracture Incidence in Middle-Aged HIV-Infected and HIV-Uninfected Women: Updated Results From the Women's Interagency HIV Study.* J Acquir Immune Defic Syndr, 2015. **70**(1): p. 54-61.
197. Gedmintas, L., et al., *Factors associated with fractures in HIV-infected persons: which factors matter?* Osteoporos Int, 2017. **28**(1): p. 239-244.
198. Battalora, L., et al., *Low bone mineral density and risk of incident fracture in HIV-infected adults.* Antivir Ther, 2016. **21**(1): p. 45-54.
199. Lang, S., et al., *HIV replication and immune status are independent predictors of the risk of myocardial infarction in HIV-infected individuals.* Clin Infect Dis, 2012. **55**(4): p. 600-7.
200. Bucher, H.C., et al., *Small dense lipoproteins, apolipoprotein B, and risk of coronary events in HIV-infected patients on antiretroviral therapy: the Swiss HIV Cohort Study.* J Acquir Immune Defic Syndr, 2012. **60**(2): p. 135-42.
201. Phillips, A.N., et al., *Interruption of antiretroviral therapy and risk of cardiovascular disease in persons with HIV-1 infection: exploratory analyses from the SMART trial.* Antivir Ther, 2008. **13**(2): p. 177-87.
202. Friis-Moller, N., et al., *Class of antiretroviral drugs and the risk of myocardial infarction.* N Engl J Med, 2007. **356**(17): p. 1723-35.
203. Triant, V.A., et al., *Association of immunologic and virologic factors with myocardial infarction rates in a US healthcare system.* J Acquir Immune Defic Syndr, 2010. **55**(5): p. 615-9.
204. Carrieri, M.P., et al., *Impact of immunodepression and moderate alcohol consumption on coronary and other arterial disease events in an 11-year cohort of HIV-infected patients on antiretroviral therapy.* BMJ Open, 2012. **2**(6).
205. Chow, F.C., et al., *Comparison of ischemic stroke incidence in HIV-infected and non-HIV-infected patients in a US health care system.* J Acquir Immune Defic Syndr, 2012. **60**(4): p. 351-8.
206. Rasmussen, L.D., et al., *Risk of cerebrovascular events in persons with and without HIV: a Danish nationwide population-based cohort study.* AIDS, 2011. **25**(13): p. 1637-46.
207. Freiberg, M.S., et al., *HIV infection and the risk of acute myocardial infarction.* JAMA Intern Med, 2013. **173**(8): p. 614-22.
208. Helleberg, M., et al., *CD4 Decline Is Associated With Increased Risk of Cardiovascular Disease, Cancer, and Death in Virally Suppressed Patients With HIV.* Clin Infect Dis, 2013.
209. van Lelyveld, S.F., et al., *Long-term complications in patients with poor immunological recovery despite virological successful HAART in Dutch ATHENA cohort.* AIDS, 2012. **26**(4): p. 465-74.
210. Marin, B., et al., *Non-AIDS-defining deaths and immunodeficiency in the era of combination antiretroviral therapy.* AIDS, 2009. **23**(13): p. 1743-53.
211. Sabin, C.A., et al., *Associations between immune depression and cardiovascular events in hiv infection.* AIDS, 2013.
212. Aberg, J.A., *Cardiovascular complications in HIV management: past, present, and future.* J Acquir Immune Defic Syndr, 2009. **50**(1): p. 54-64.
213. Hippisley-Cox, J., C. Coupland, and P. Brindle, *Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study.* BMJ, 2017. **357**: p. j2099.
214. Currier, J.S., *Update on cardiovascular complications in HIV infection.* Top HIV Med, 2009. **17**(3): p. 98-103.

215. Friis-Moller, N., et al., *Predicting the risk of cardiovascular disease in HIV-infected patients: the data collection on adverse effects of anti-HIV drugs study.* Eur J Cardiovasc Prev Rehabil, 2010. **17**(5): p. 491-501.
216. Law, M.G., et al., *The use of the Framingham equation to predict myocardial infarctions in HIV-infected patients: comparison with observed events in the D:A:D Study.* HIV Med, 2006. **7**(4): p. 218-30.
217. Elion, R.A., et al., *Recent Abacavir Use Increases Risk of Type 1 and Type 2 Myocardial Infarctions Among Adults With HIV.* J Acquir Immune Defic Syndr, 2018. **78**(1): p. 62-72.
218. Feinstein, M.J., J.A. Delaney, and H.M. Crane, *Cardiovascular Risk Prediction in Patients With Human Immunodeficiency Virus-Reply.* JAMA Cardiol, 2017. **2**(9): p. 1048-1049.
219. Friis-Moller, N., et al., *An updated prediction model of the global risk of cardiovascular disease in HIV-positive persons: The Data-collection on Adverse Effects of Anti-HIV Drugs (D:A:D) study.* Eur J Prev Cardiol, 2016. **23**(2): p. 214-23.
220. Silverberg, M.J., et al., *Immunodeficiency and risk of myocardial infarction among HIV-positive individuals with access to care.* J Acquir Immune Defic Syndr, 2014. **65**(2): p. 160-6.
221. Salinas, J.L., et al., *Baseline, Time-Updated, and Cumulative HIV Care Metrics for Predicting Acute Myocardial Infarction and All-Cause Mortality.* Clin Infect Dis, 2016. **63**(11): p. 1423-1430.
222. Capeau, J., et al., *Ten-year diabetes incidence in 1046 HIV-infected patients started on a combination antiretroviral treatment.* AIDS, 2012. **26**(3): p. 303-14.
223. Justman, J.E., et al., *Protease inhibitor use and the incidence of diabetes mellitus in a large cohort of HIV-infected women.* J Acquir Immune Defic Syndr, 2003. **32**(3): p. 298-302.
224. Ledergerber, B., et al., *Factors associated with the incidence of type 2 diabetes mellitus in HIV-infected participants in the Swiss HIV Cohort Study.* Clin Infect Dis, 2007. **45**(1): p. 111-9.
225. Rotger, M., et al., *Impact of single nucleotide polymorphisms and of clinical risk factors on new-onset diabetes mellitus in HIV-infected individuals.* Clin Infect Dis, 2010. **51**(9): p. 1090-8.
226. Petoumenos, K., et al., *Predicting the short-term risk of diabetes in HIV-positive patients: the Data Collection on Adverse Events of Anti-HIV Drugs (D:A:D) study.* J Int AIDS Soc, 2012. **15**(2): p. 17426.
227. Hippisley-Cox, J. and C. Coupland, *Development and validation of QDiabetes-2018 risk prediction algorithm to estimate future risk of type 2 diabetes: cohort study.* BMJ, 2017. **359**: p. j5019.
228. Butt, A.A., et al., *Risk of diabetes in HIV infected veterans pre- and post-HAART and the role of HCV coinfection.* Hepatology, 2004. **40**(1): p. 115-9.
229. Spagnuolo, V., et al., *Associations of statins and antiretroviral drugs with the onset of type 2 diabetes among HIV-1-infected patients.* BMC Infect Dis, 2017. **17**(1): p. 43.
230. Brambilla, A.M., et al., *Stavudine or indinavir-containing regimens are associated with an increased risk of diabetes mellitus in HIV-infected individuals.* AIDS, 2003. **17**(13): p. 1993-5.
231. Capeau, J., et al., *Ten-year diabetes incidence in 1046 HIV-infected patients started on a combination antiretroviral treatment.* AIDS, 2012. **26**(3): p. 303-14.
232. Tripathi, A., et al., *Incidence of diabetes mellitus in a population-based cohort of HIV-infected and non-HIV-infected persons: the impact of clinical and therapeutic factors over time.* Diabet Med, 2014. **31**(10): p. 1185-93.

233. De Luca, A., et al., *Active HCV Replication but Not HCV or CMV Seropositive Status Is Associated With Incident and Prevalent Type 2 Diabetes in Persons Living With HIV.* J Acquir Immune Defic Syndr, 2017. **75**(4): p. 465-471.
234. Seng, R., et al., *Rapid CD4+ cell decrease after transient cART initiated during primary HIV infection (ANRS PRIMO and SEROCO cohorts).* J Acquir Immune Defic Syndr, 2008. **49**(3): p. 251-8.
235. D:A:D. *Participating Cohorts*; Available from: https://www.chip.dk/Studies/DAD/Study-Group.
236. Silverberg, M.J., et al., *Risk of anal cancer in HIV-infected and HIV-uninfected individuals in North America.* Clin Infect Dis, 2012. **54**(7): p. 1026-34.
237. Wong, C., et al., *First Occurrence of Diabetes, Chronic Kidney Disease, and Hypertension Among North American HIV-Infected Adults, 2000-2013.* Clin Infect Dis, 2017. **64**(4): p. 459-467.
238. D:A:D. *Manual of Operations* January 2013; Available from: https://www.chip.dk/Portals/0/files/Study%20documents/DAD_MOOP_revised2013.pdf.
239. EuroSIDA. *EuroSIDA list of definitions* Available from: http://www.cphiv.dk/Portals/0/ES_List%20of%20Definitions%20on%20Clinical%20events%20FU%20d42%20Summer%202015_2015JUN03.pdf.
240. Swiss HIV Cohort Study, *Non-AIDS defining events* Available from: http://www.shcs.ch/126-8-non-aids-defining-events.
241. *American College of Cardiology*; Available from: www.acc.org.
242. *American Diabetes Association*; Available from: http://www.diabetes.org/.
243. *British Society of Gastroenterology*; Available from: https://www.bsg.org.uk/.
244. *American College of Gastroenterology*; Available from: https://gi.org/.
245. *World Health Organisation*; Available from: http://www.who.int/en/.
246. *American Heart Association*; Available from: http://www.heart.org/HEARTORG/.
247. *American Stroke Association*; Available from: http://www.strokeassociation.org/STROKEORG/.
248. *European Renal Assocation*; Available from: http://web.era-edta.org/.
249. *American Society of Nephrology*; Available from: https://www.asn-online.org/.
250. *International Osteoporosis Foundation*; Available from: https://www.iofbonehealth.org/.
251. *European Society for Medical Oncology*; Available from: http://www.esmo.org/.
252. *American Society of Clinical Oncology*; Available from: https://www.asco.org/.
253. *National Institutes of Health*; Available from: https://www.nih.gov/.
254. Luepker, R.V., et al., *Case definitions for acute coronary heart disease in epidemiology and clinical research studies: a statement from the AHA Council on Epidemiology and Prevention; AHA Statistics Committee; World Heart Federation Council on Epidemiology and Prevention; the European Society of Cardiology Working Group on Epidemiology and Prevention; Centers for Disease Control and Prevention; and the National Heart, Lung, and Blood Institute.* Circulation, 2003. **108**(20): p. 2543-9.
255. Thygesen, K., et al., *Third universal definition of myocardial infarction.* Circulation, 2012. **126**(16): p. 2020-35.
256. *The HIV Cohorts Data Exchange Protocol (HICDEP)*; Available from: http://www.hicdep.org/.
257. *Concerted Action on SeroConversion to AIDS and Death in Europe*; Available from: http://www.ctu.mrc.ac.uk/cascade/about_cascade.asp.

258. Babiker, A., et al., *Changes over calendar time in the risk of specific first AIDS-defining events following HIV seroconversion, adjusting for competing risks.* Int J Epidemiol, 2002. **31**(5): p. 951-8.
259. Pantazis, N., et al., *The effect of antiretroviral treatment of different durations in primary HIV infection.* AIDS, 2008. **22**(18): p. 2441-50.
260. Pantazis, N., et al., *Temporal trends in prognostic markers of HIV-1 virulence and transmissibility: an observational cohort study.* Lancet HIV, 2014. **1**(3): p. e119-26.
261. Touloumi, G., et al., *Virologic and immunologic response to cART by HIV-1 subtype in the CASCADE collaboration.* PLoS One, 2013. **8**(7): p. e71174.
262. Touloumi, G., et al., *Impact of HIV-1 subtype on CD4 count at HIV seroconversion, rate of decline, and viral load set point in European seroconverter cohorts.* Clin Infect Dis, 2013. **56**(6): p. 888-97.
263. 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.
264. van Santen, D.K., et al., *Lack of decline in hepatitis C virus incidence among HIV-positive men who have sex with men during 1990-2014.* J Hepatol, 2017. **67**(2): p. 255-262.
265. Pantazis, N., et al., *Differences in HIV natural history among African and non-African seroconverters in Europe and seroconverters in sub-Saharan Africa.* PLoS One, 2012. **7**(3): p. e32369.
266. Huang, X., et al., *Rate of CD4 decline and HIV-RNA change following HIV seroconversion in men who have sex with men: a comparison between the Beijing PRIMO and CASCADE cohorts.* J Acquir Immune Defic Syndr, 2013. **62**(4): p. 441-6.
267. Simmons, R., et al., *HIV Incidence Estimates Using the Limiting Antigen Avidity EIA Assay at Testing Sites in Kiev City, Ukraine: 2013-2014.* PLoS One, 2016. **11**(6): p. e0157179.
268. Simmons, R., et al., *HIV Testing and Diagnosis Rates in Kiev, Ukraine: April 2013-March 2014.* PLoS One, 2015. **10**(8): p. e0137062.
269. Olson, A.D., et al., *An evaluation of HIV elite controller definitions within a large seroconverter cohort collaboration.* PLoS One, 2014. **9**(1): p. e86719.
270. van der Helm, J.J., et al., *Characterisation of long-term non-progression of HIV-1 infection after seroconversion: a cohort study.* Lancet HIV, 2014. **1**(1): p. e41-8.
271. Jarrin, I., et al., *Does rapid HIV disease progression prior to combination antiretroviral therapy hinder optimal CD4+ T-cell recovery once HIV-1 suppression is achieved?* AIDS, 2015. **29**(17): p. 2323-33.
272. Olson, A.D., et al., *Evaluation of rapid progressors in HIV infection as an extreme phenotype.* J Acquir Immune Defic Syndr, 2014. **67**(1): p. 15-21.
273. Olson, A., et al., *Temporal trends of transmitted HIV drug resistance in a multinational seroconversion cohort.* AIDS, 2017.
274. Stirrup, O.T., et al., *Predictors of CD4 cell recovery following initiation of antiretroviral therapy among HIV-1 positive patients with well-estimated dates of seroconversion.* HIV Med, 2017.
275. Fearnhill, E., et al., *A Phylogenetic Analysis of HIV-1 Sequences in Kiev: Findings among Key Populations.* Clin Infect Dis, 2017.
276. EuroCoord, *Founding networks*; Available from: http://www.eurocoord.net/partners/founding_networks.
277. Huang, X., et al., *Precise determination of time to reach viral load set point after acute HIV-1 infection.* J Acquir Immune Defic Syndr, 2012. **61**(4): p. 448-54.

278. Vittinghoff, E. and C.E. McCulloch, *Relaxing the rule of ten events per variable in logistic and Cox regression.* Am J Epidemiol, 2007. **165**(6): p. 710-8.
279. Tai, B.C. and D. Machin, *Regression methods for medical research*. 2013: Wiley Blackwell.
280. Royston, P., G. Ambler, and W. Sauerbrei, *The use of fractional polynomials to model continuous risk variables in epidemiology.* Int J Epidemiol, 1999. **28**(5): p. 964-74.
281. Royston, P. and W. Sauerbrei, *Multivariable Model - Building : A Pragmatic Approach to Regression Anaylsis based on Fractional Polynomials for Modelling Continuous Variables*. 2008: John Wiley and Sons Ltd.
282. Mazumdar, M. and J.R. Glassman, *Categorizing a prognostic variable: review of methods, code for easy implementation and applications to decision-making about cancer treatments.* Stat Med, 2000. **19**(1): p. 113-32.
283. Bilinger, M. and W.G. Manning, *Measuring overfitting in nonlinear models: A new method and an application to health expenditures.* Health Economics, 2015. **24(1)**: p. 75-85.
284. Baneshi, M.R., F. Nakhaee, and M. Law, *On the Use of Fractional Polynomial Models to Assess Preventive Aspect of Variables: An Example in Prevention of Mortality Following HIV Infection.* Int J Prev Med, 2013. **4**(4): p. 414-9.
285. Crane, H.M., et al., *Lessons learned from the design and implementation of myocardial infarction adjudication tailored for HIV clinical cohorts.* Am J Epidemiol, 2014. **179**(8): p. 996-1005.
286. Viera, A.J. and J.M. Garrett, *Understanding interobserver agreement: the kappa statistic.* Fam Med, 2005. **37**(5): p. 360-3.
287. *The UK register of HIV seroconverters: methods and analytical issues. UK register of HIV seroconverters (UKRHS) Steering Committee.* Epidemiol Infect, 1996. **117**(2): p. 305-12.
288. *HIV in the United Kingdom: 2013 Report*. 2013, Public Health England.
289. Mary-Krause, M., et al., *Cohort Profile: French hospital database on HIV (FHDH-ANRS CO4).* Int J Epidemiol, 2014. **43**(5): p. 1425-36.
290. Amorosa, V. and P. Tebas, *Bone disease and HIV infection.* Clin Infect Dis, 2006. **42**(1): p. 108-14.
291. Bedimo, R., N.M. Maalouf, and V. Lo Re, 3rd, *Hepatitis C virus coinfection as a risk factor for osteoporosis and fracture.* Curr Opin HIV AIDS, 2016. **11**(3): p. 285-93.
292. Smurzynski, M., et al., *AIDS clinical trials group longitudinal linked randomized trials (ALLRT): rationale, design, and baseline characteristics.* HIV Clin Trials, 2008. **9**(4): p. 269-82.
293. Adimora, A.A., et al., *Cohort Profile: The Women's Interagency HIV Study (WIHS).* Int J Epidemiol, 2018. **47**(2): p. 393-394i.
294. Vellozzi, C., et al., *The study to understand the natural history of HIV and AIDS in the era of effective therapy (SUN Study).* Am J Epidemiol, 2009. **169**(5): p. 642-52.
295. Delgado-Rodriguez, M. and J. Llorca, *Bias.* J Epidemiol Community Health, 2004. **58**(8): p. 635-41.
296. Obel, N., et al., *Cohort profile: the Danish HIV cohort study.* Int J Epidemiol, 2009. **38**(5): p. 1202-6.
297. Slomka, J., et al., *Decisions to participate in research: views of underserved minority drug users with or at risk for HIV.* AIDS Care, 2008. **20**(10): p. 1224-32.
298. Robbins, G.K., et al., *Comparison of sequential three-drug regimens as initial therapy for HIV-1 infection.* N Engl J Med, 2003. **349**(24): p. 2293-303.
299. Hammer, S.M., et al., *Dual vs single protease inhibitor therapy following antiretroviral treatment failure: a randomized trial.* JAMA, 2002. **288**(2): p. 169-80.

300. Hessol, N.A., et al., *Retention of women enrolled in a prospective study of human immunodeficiency virus infection: impact of race, unstable housing, and use of human immunodeficiency virus therapy.* Am J Epidemiol, 2001. **154**(6): p. 563-73.
301. José M. Zuniga, A.W., Amin Ghaziani, John G. Bartlett, *A Decade of HAART: The Development and Global Impact of Highly Active Antiretroviral Therapy*. 2008: Oxford University Press.
302. Sinclair, G.D., *Homosexuality and the military: a review of the literature.* J Homosex, 2009. **56**(6): p. 701-18.
303. Sudre, P., et al., *Clinical epidemiology and research on HIV infection in Switzerland: the Swiss HIV Cohort Study 1988-2000.* Schweiz Med Wochenschr, 2000. **130**(41): p. 1493-500.
304. Guy, R.J., et al., *HIV diagnoses in Australia: Diverging epidemics within a low-prevalence country.* Medical Journal of Australia, 2007. **187**(8): p. 437-440.
305. McDonald, A.M., et al., *The pattern of diagnosed HIV infection in Australia, 1984-1992.* AIDS, 1994. **8**(4): p. 513-9.
306. Omland, L.H., M.G. Ahlstrom, and N. Obel, *Cohort profile update: the Danish HIV Cohort Study (DHCS).* Int J Epidemiol, 2014. **43**(6): p. 1769-9e.
307. Brady, A.P., *Error and discrepancy in radiology: inevitable or avoidable?* Insights Imaging, 2017. **8**(1): p. 171-182.
308. Reid, D.C., et al., *Etiology and clinical course of missed spine fractures.* J Trauma, 1987. **27**(9): p. 980-6.
309. Klotzbuecher, C.M., et al., *Patients with prior fractures have an increased risk of future fractures: a summary of the literature and statistical synthesis.* J Bone Miner Res, 2000. **15**(4): p. 721-39.
310. Ismail, A.A., et al., *Validity of self-report of fractures: results from a prospective study in men and women across Europe. EPOS Study Group. European Prospective Osteoporosis Study Group.* Osteoporos Int, 2000. **11**(3): p. 248-54.
311. Ivers, R.Q., et al., *The accuracy of self-reported fractures in older people.* J Clin Epidemiol, 2002. **55**(5): p. 452-7.
312. Rothman, K.J., S. Greenland, and T.L. Lash, *Modern Epidemiology*. 3rd mid-cycle revised edition ed. 2013, Philadelphia, United States: Lippincott Williams and Wilkins.
313. Gange, S.J., et al., *Cohort profile: the North American AIDS Cohort Collaboration on Research and Design (NA-ACCORD).* Int J Epidemiol, 2007. **36**(2): p. 294-301.
314. National Clinical Guidelines Centre (UK), *Osteoporosis: Fragility Fracture Risk: Assessing the Risk of Fragility Fracture: Risk assessment tools (FRAX, QFracture, BMD)* NICE Clinical Guidelines. Vol. No. 146. 2012, London: Royal College of Physicians (UK)
315. Battalora, L.A., B. Young, and E.T. Overton, *Bones, Fractures, Antiretroviral Therapy and HIV.* Curr Infect Dis Rep, 2014. **16**(2): p. 393.
316. Teichmann, J., et al., *Osteopenia in HIV-infected women prior to highly active antiretroviral therapy.* J Infect, 2003. **46**(4): p. 221-7.
317. Fakruddin, J.M. and J. Laurence, *HIV envelope gp120-mediated regulation of osteoclastogenesis via receptor activator of nuclear factor kappa B ligand (RANKL) secretion and its modulation by certain HIV protease inhibitors through interferon-gamma/RANKL cross-talk.* J Biol Chem, 2003. **278**(48): p. 48251-8.
318. Fakruddin, J.M. and J. Laurence, *HIV-1 Vpr enhances production of receptor of activated NF-kappaB ligand (RANKL) via potentiation of glucocorticoid receptor activity.* Arch Virol, 2005. **150**(1): p. 67-78.
319. Serrano, S., et al., *Bone remodelling in human immunodeficiency virus-1-infected patients. A histomorphometric study.* Bone, 1995. **16**(2): p. 185-91.

320. Hernan, M.A., et al., *Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology.* Am J Epidemiol, 2002. **155**(2): p. 176-84.
321. Swiss HIV Cohort Study, et al., *Cohort profile: the Swiss HIV Cohort study.* Int J Epidemiol, 2010. **39**(5): p. 1179-89.
322. Laribi, S., et al., *Trends in death attributed to myocardial infarction, heart failure and pulmonary embolism in Europe and Canada over the last decade.* Qjm-an International Journal of Medicine, 2014. **107**(10): p. 813-820.
323. Post, W.S., *Predicting and preventing cardiovascular disease in HIV-infected patients.* Top Antivir Med, 2011. **19**(5): p. 169-73.
324. Barbaro, G., *Heart and HAART: Two sides of the coin for HIV-associated cardiology issues.* World J Cardiol, 2010. **2**(3): p. 53-7.
325. Gibellini, D., et al., *HIV-related mechanisms in atherosclerosis and cardiovascular diseases.* J Cardiovasc Med (Hagerstown), 2013.
326. Stein, J.H., *Cardiovascular risk and dyslipidemia management in HIV-infected patients.* Top Antivir Med, 2012. **20**(4): p. 129-33; quiz 123-4.
327. Thygesen, K., et al., *Fourth universal definition of myocardial infarction (2018).* Eur Heart J, 2018.
328. Alpert, J.S., et al., *Myocardial infarction redefined--a consensus document of The Joint European Society of Cardiology/American College of Cardiology Committee for the redefinition of myocardial infarction.* J Am Coll Cardiol, 2000. **36**(3): p. 959-69.
329. McCormick, N., et al., *Validity of myocardial infarction diagnoses in administrative databases: a systematic review.* PLoS One, 2014. **9**(3): p. e92286.
330. Tunstall-Pedoe, H., et al., *Estimation of contribution of changes in coronary care to improving survival, event rates, and coronary heart disease mortality across the WHO MONICA Project populations.* Lancet, 2000. **355**(9205): p. 688-700.
331. Tunstall-Pedoe, H., et al., *Myocardial infarction and coronary deaths in the World Health Organization MONICA Project. Registration procedures, event rates, and case-fatality rates in 38 populations from 21 countries in four continents.* Circulation, 1994. **90**(1): p. 583-612.
332. World Health Organization, *Myocardial Infarction Community Registers*, in *Public Health in Europe*. 1976, Regional Office for Europe. Copenhagen WHO.
333. Mendis, S., et al., *World Health Organization definition of myocardial infarction: 2008-09 revision.* Int J Epidemiol, 2011. **40**(1): p. 139-46.
334. Mendis, S., et al., *World Health Organization definition of myocardial infarction: 2008-09 revision.* Int J Epidemiol, 2011. **40**(1): p. 139-46.
335. Collaborators, G.B.D.T., *Smoking prevalence and attributable disease burden in 195 countries and territories, 1990-2015: a systematic analysis from the Global Burden of Disease Study 2015.* Lancet, 2017. **389**(10082): p. 1885-1906.
336. Lakier, J.B., *Smoking and cardiovascular disease.* Am J Med, 1992. **93**(1A): p. 8S-12S.
337. Rigotti, N.A. and R.C. Pasternak, *Cigarette smoking and coronary heart disease: risks and management.* Cardiol Clin, 1996. **14**(1): p. 51-68.
338. Ferguson, J.L., et al., *Myocardial infarction redefined: the new ACC/ESC definition, based on cardiac troponin, increases the apparent incidence of infarction.* Heart, 2002. **88**(4): p. 343-7.
339. Crane, H.M., et al., *Types of Myocardial Infarction Among Human Immunodeficiency Virus-Infected Individuals in the United States.* JAMA Cardiol, 2017. **2**(3): p. 260-267.
340. Kusnik-Joinville, O., et al., *Prevalence and treatment of diabetes in France: trends between 2000 and 2005.* Diabetes Metab, 2008. **34**(3): p. 266-72.

341. Mulnier, H.E., et al., *Mortality in people with type 2 diabetes in the UK.* Diabet Med, 2006. **23**(5): p. 516-21.
342. Araujo, S., et al., *Prevalence of insulin resistance and risk of diabetes mellitus in HIV-infected patients receiving current antiretroviral drugs.* Eur J Endocrinol, 2014. **171**(5): p. 545-54.
343. Thomas, C.C. and L.H. Philipson, *Update on diabetes classification.* Med Clin North Am, 2015. **99**(1): p. 1-16.
344. Melmer, A. and M. Laimer, *Treatment Goals in Diabetes.* Endocr Dev, 2016. **31**: p. 1-27.
345. Paik, I.J. and D.P. Kotler, *The prevalence and pathogenesis of diabetes mellitus in treated HIV-infection.* Best Pract Res Clin Endocrinol Metab, 2011. **25**(3): p. 469-78.
346. Dwyer-Lindgren, L., et al., *Diagnosed and Undiagnosed Diabetes Prevalence by County in the U.S., 1999-2012.* Diabetes Care, 2016. **39**(9): p. 1556-62.
347. Leong, A., et al., *Estimating the population prevalence of diagnosed and undiagnosed diabetes.* Diabetes Care, 2013. **36**(10): p. 3002-8.
348. Bonaldi, C., et al., *A first national prevalence estimate of diagnosed and undiagnosed diabetes in France in 18- to 74-year-old individuals: the French Nutrition and Health Survey 2006/2007.* Diabet Med, 2011. **28**(5): p. 583-9.
349. Fletcher, B., M. Gulanick, and C. Lamendola, *Risk factors for type 2 diabetes mellitus.* J Cardiovasc Nurs, 2002. **16**(2): p. 17-23.
350. Arora, S., et al., *HIV and bone loss.* Curr Osteoporos Rep, 2010. **8**(4): p. 219-26.
351. Tousoulis, D., M. Charakida, and C. Stefanadis, *Endothelial function and inflammation in coronary artery disease.* Heart, 2006. **92**(4): p. 441-4.
352. Olie, R.H., P.E.J. van der Meijden, and H. Ten Cate, *The coagulation system in atherothrombosis: Implications for new therapeutic strategies.* Res Pract Thromb Haemost, 2018. **2**(2): p. 188-198.
353. Hunt, P.W., S.A. Lee, and M.J. Siedner, *Immunologic Biomarkers, Morbidity, and Mortality in Treated HIV Infection.* J Infect Dis, 2016. **214 Suppl 2**: p. S44-50.
354. Hsu, D.C. and I. Sereti, *Serious Non-AIDS Events: Therapeutic Targets of Immune Activation and Chronic Inflammation in HIV Infection.* Drugs, 2016. **76**(5): p. 533-49.
355. Saracci, R., *Survival-related biases survive well.* Int J Epidemiol, 2007. **36**(1): p. 244-6.
356. Lesko, C.R., et al., *Collaborative, pooled and harmonized study designs for epidemiologic research: challenges and opportunities.* Int J Epidemiol, 2018. **47**(2): p. 654-668.
357. *Meta-Analyses of Aggregate Data or Individual Participant Data Meta-Analyses (Retrospectively/Prospectively Pooled Analyses)*. 23/07/2018]; Available from: https://www.mailman.columbia.edu/research/population-health-methods/meta-analyses-aggregate-data-or-individual-participant-data-meta-analyses-retrospectively-and.
358. Rothman, K.J., *Modern Epidemiology* Third (Revised) ed. 2013, Philadelphia, United States: Lippincott Williams and Wilkins. 816.
359. Weintraub, W.S., T.F. Luscher, and S. Pocock, *The perils of surrogate endpoints.* Eur Heart J, 2015. **36**(33): p. 2212-8.
360. Weintraub, W.S., *Statistical Approaches to Composite Endpoints.* JACC Cardiovasc Interv, 2016. **9**(22): p. 2289-2291.
361. Madeddu, G., et al., *CD4 cell count and the risk of infective and non-infective serious non-AIDS events in HIV-positive persons seen far care in Italy* in *HIV Drug Therapy Glasgow Congress 2014*. 2014: Glasgow, UK.

362. Serrano-Villar, S., et al., *Increased risk of serious non-AIDS-related events in HIV-infected subjects on antiretroviral therapy associated with a low CD4/CD8 ratio.* PLoS One, 2014. **9**(1): p. e85798.
363. Althubaiti, A., *Information bias in health research: definition, pitfalls, and adjustment methods.* J Multidiscip Healthc, 2016. **9**: p. 211-7.
364. Atkins, D., et al., *Grading quality of evidence and strength of recommendations.* BMJ, 2004. **328**(7454): p. 1490.
365. Connor Gorber, S., et al., *The accuracy of self-reported smoking: a systematic review of the relationship between self-reported and cotinine-assessed smoking status.* Nicotine Tob Res, 2009. **11**(1): p. 12-24.
366. Sainani, K., *The limitations of statistical adjustment.* PM R, 2011. **3**(9): p. 868-72.
367. Lubin, J.H., et al., *Risk of Cardiovascular Disease from Cumulative Cigarette Use and the Impact of Smoking Intensity.* Epidemiology, 2016. **27**(3): p. 395-404.
368. Frazier, E.L., et al., *Trends in cigarette smoking among adults with HIV compared with the general adult population, United States - 2009-2014.* Prev Med, 2018. **111**: p. 231-234.
369. Marcon, A., et al., *Trends in smoking initiation in Europe over 40 years: A retrospective cohort study.* PLoS One, 2018. **13**(8): p. e0201881.
370. Tron, L., et al., *Tobacco smoking in HIV-infected versus general population in france: heterogeneity across the various groups of people living with HIV.* PLoS One, 2014. **9**(9): p. e107451.
371. Jakobsen, J.C., et al., *Do direct acting antivirals cure chronic hepatitis C?* BMJ, 2018. **361**: p. k1382.
372. Kanis, J.A., et al., *Development and use of FRAX in osteoporosis.* Osteoporos Int, 2010. **21 Suppl 2**: p. S407-13.
373. Sopori, M.L. and W. Kozak, *Immunomodulatory effects of cigarette smoke.* J Neuroimmunol, 1998. **83**(1-2): p. 148-56.
374. Kasang, C., et al., *Effects of Prednisolone on Disease Progression in Antiretroviral-Untreated HIV Infection: A 2-Year Randomized, Double-Blind Placebo-Controlled Clinical Trial.* PLoS One, 2016. **11**(1): p. e0146678.
375. Johnson, K.D., et al., *Longitudinal association between BMI at diagnosis and HIV disease progression.* AIDS Behav, 2014. **18**(11): p. 2249-57.
376. Baum, M.K., et al., *Alcohol use accelerates HIV disease progression.* AIDS Res Hum Retroviruses, 2010. **26**(5): p. 511-8.
377. Thompson, P.D., *CHD risk equivalents and the metabolic syndrome. Trial evidence supports aggressive management.* Postgrad Med, 2003. **114**(2 Suppl): p. 22-30.
378. Briasoulis, A. and G.L. Bakris, *Chronic kidney disease as a coronary artery disease risk equivalent.* Curr Cardiol Rep, 2013. **15**(3): p. 340.
379. Kristman, V., M. Manno, and P. Cote, *Loss to follow-up in cohort studies: how much is too much?* Eur J Epidemiol, 2004. **19**(8): p. 751-60.
380. Howe, C.J., et al., *Selection Bias Due to Loss to Follow Up in Cohort Studies.* Epidemiology, 2016. **27**(1): p. 91-7.
381. Sedgwick, P., *Bias in observational study designs: prospective cohort studies.* BMJ, 2014. **349**: p. g7731.
382. Altman, D.G., *Statistics in medical journals: some recent trends.* Stat Med, 2000. **19**(23): p. 3275-89.
383. Jackson, D., et al., *Relaxing the independent censoring assumption in the Cox proportional hazards model using multiple imputation.* Stat Med, 2014. **33**(27): p. 4681-94.
384. Crainiceanu, C.M. and A.J. Goldsmith, *Bayesian Functional Data Analysis Using WinBUGS.* J Stat Softw, 2010. **32**(11).

385. Howe, C.J., et al., *Limitation of inverse probability-of-censoring weights in estimating survival in the presence of strong selection bias.* Am J Epidemiol, 2011. **173**(5): p. 569-77.
386. Bagkeris, E., et al., *Cohort profile: The Pharmacokinetic and clinical Observations in PeoPle over fiftY (POPPY) study.* Int J Epidemiol, 2018.
387. Justice, A.C., et al., *Veterans Aging Cohort Study (VACS): Overview and description.* Med Care, 2006. **44**(8 Suppl 2): p. S13-24.
388. Wolbers, M., et al., *Delayed diagnosis of HIV infection and late initiation of antiretroviral therapy in the Swiss HIV Cohort Study.* Hiv Medicine, 2008. **9**(6): p. 397-405.
389. Stohr, W., et al., *CD4 cell count and initiation of antiretroviral therapy: trends in seven UK centres, 1997-2003.* HIV Med, 2007. **8**(3): p. 135-41.
390. *Institute of Medicine (US) Committee on Technological Innovation in Medicine: Modern Methods of Clinical Investigation: Medical Innovation at the Crossroads. 5: Strengths and Weaknesses of Health Insurance Data Systems for Assessing Outcomes.* , ed. A. Gelijns. Vol. I. 1990, Washington (DC): National Academies Press (US);.
391. Chene, G., et al., *Cohort Profile: Collaboration of Observational HIV Epidemiological Research Europe (COHERE) in EuroCoord.* Int J Epidemiol, 2017. **46**(3): p. 797-797n.
392. European Centre for Disease Control, *Improving HIV Data Comparability in Migrant Populations and Ethnic Minorities in EU/EEA/EFTA Countries: Findings from a Literature Review and Expert Panel .Solna, Sweden: European Centre for Disease Control.* 2011.
393. Mussini, C., et al., *CD4/CD8 ratio normalisation and non-AIDS-related events in individuals with HIV who achieve viral load suppression with antiretroviral therapy: an observational cohort study.* Lancet HIV, 2015. **2**(3): p. e98-106.
394. Freemantle, N., et al., *Making inferences on treatment effects from real world data: propensity scores, confounding by indication, and other perils for the unwary in observational research.* BMJ, 2013. **347**: p. f6409.
395. Petri, H. and J. Urquhart, *Channeling bias in the interpretation of drug effects.* Stat Med, 1991. **10**(4): p. 577-81.
396. Salas, M., A. Hofman, and B.H. Stricker, *Confounding by indication: an example of variation in the use of epidemiologic terminology.* Am J Epidemiol, 1999. **149**(11): p. 981-3.
397. Sterne, J., *G-estimation of causal effects, allowing for time-varying confounding.* The Stata Journal 2002.
398. Shrier, I. and R.W. Platt, *Reducing bias through directed acyclic graphs.* BMC Med Res Methodol, 2008. **8**: p. 70.
399. Cole, S.R., et al., *Illustrating bias due to conditioning on a collider.* Int J Epidemiol, 2010. **39**(2): p. 417-20.
400. Copeland, K.T., et al., *Bias due to misclassification in the estimation of relative risk.* Am J Epidemiol, 1977. **105**(5): p. 488-95.
401. Lambert, J., *Statistics in brief: how to assess bias in clinical studies?* Clin Orthop Relat Res, 2011. **469**(6): p. 1794-6.
402. Bruyninckx, R., et al., *Signs and symptoms in diagnosing acute myocardial infarction and acute coronary syndrome: a diagnostic meta-analysis.* Br J Gen Pract, 2008. **58**(547): p. 105-11.
403. Mattila, K.J., et al., *Role of infection as a risk factor for atherosclerosis, myocardial infarction, and stroke.* Clin Infect Dis, 1998. **26**(3): p. 719-34.
404. Libby, P., et al., *Inflammation in atherosclerosis: transition from theory to practice.* Circ J, 2010. **74**(2): p. 213-20.

405. McCarthy, C.P., M. Vaduganathan, and J.L. Januzzi, Jr., *Type 2 Myocardial Infarction-Diagnosis, Prognosis, and Treatment.* JAMA, 2018. **320**(5): p. 433-434.
406. McCarthy, C.P., J.L. Januzzi, Jr., and H.K. Gaggin, *Type 2 Myocardial Infarction- An Evolving Entity.* Circ J, 2018. **82**(2): p. 309-315.
407. Gupta, S., et al., *Type 2 versus type 1 myocardial infarction: a comparison of clinical characteristics and outcomes with a meta-analysis of observational studies.* Cardiovasc Diagn Ther, 2017. **7**(4): p. 348-358.
408. Farr, J.N., et al., *Fracture Incidence and Characteristics in Young Adults Aged 18 to 49 Years: A Population-Based Study.* J Bone Miner Res, 2017. **32**(12): p. 2347-2354.
409. Rewers, M. and J. Ludvigsson, *Environmental risk factors for type 1 diabetes.* Lancet, 2016. **387**(10035): p. 2340-2348.
410. Pajak, A., et al., *Geographical variation in the major risk factors of coronary heart disease in men and women aged 35-64 years*. 1988, WHO MONICA Project.
411. Hajar, R., *Evolution of Myocardial Infarction and its Biomarkers: A Historical Perspective.* Heart Views, 2016. **17**(4): p. 167-172.
412. Gebel, E., *The start of something good: the discovery of HbA(1c) and the American Diabetes Association Samuel Rahbar Outstanding Discovery Award.* Diabetes Care, 2012. **35**(12): p. 2429-31.
413. Komatsu, A., et al., *Osteoporosis-Related Fractures in HIV-Infected Patients Receiving Long-Term Tenofovir Disoproxil Fumarate: An Observational Cohort Study.* Drug Saf, 2018. **41**(9): p. 843-848.
414. Curtis, E.M., et al., *Epidemiology of fractures in the United Kingdom 1988-2012: Variation with age, sex, geography, ethnicity and socioeconomic status.* Bone, 2016. **87**: p. 19-26.
415. Hoy, J.F., et al., *Immediate Initiation of Antiretroviral Therapy for HIV Infection Accelerates Bone Loss Relative to Deferring Therapy: Findings from the START Bone Mineral Density Substudy, a Randomized Trial.* J Bone Miner Res, 2017. **32**(9): p. 1945-1955.
416. Grund, B., et al., *Continuous antiretroviral therapy decreases bone mineral density.* AIDS, 2009. **23**(12): p. 1519-29.
417. McComsey, G.A., et al., *Bone mineral density and fractures in antiretroviral-naive persons randomized to receive abacavir-lamivudine or tenofovir disoproxil fumarate-emtricitabine along with efavirenz or atazanavir-ritonavir: Aids Clinical Trials Group A5224s, a substudy of ACTG A5202.* J Infect Dis, 2011. **203**(12): p. 1791-801.
418. Stellbrink, H.J., et al., *Comparison of changes in bone density and turnover with abacavir-lamivudine versus tenofovir-emtricitabine in HIV-infected adults: 48-week results from the ASSERT study.* Clin Infect Dis, 2010. **51**(8): p. 963-72.
419. Bedimo, R., L. Rosenblatt, and J. Myers, *Systematic review of renal and bone safety of the antiretroviral regimen efavirenz, emtricitabine, and tenofovir disoproxil fumarate in patients with HIV infection.* HIV Clin Trials, 2016. **17**(6): p. 246-266.
420. Baker, J.V. and D. Duprez, *Biomarkers and HIV-associated cardiovascular disease.* Curr Opin HIV AIDS, 2010. **5**(6): p. 511-6.
421. Lowe, G.D., et al., *C-reactive protein, fibrin D-dimer, and incident ischemic heart disease in the Speedwell study: are inflammation and fibrin turnover linked in pathogenesis?* Arterioscler Thromb Vasc Biol, 2001. **21**(4): p. 603-10.
422. Sabin, C.A., et al., *Use of nucleoside reverse transcriptase inhibitors and risk of myocardial infarction in HIV-infected patients enrolled in the D:A:D study: a multi-cohort collaboration.* Lancet, 2008. **371**(9622): p. 1417-26.
423. Mary-Krause, M., et al., *Increased risk of myocardial infarction with duration of protease inhibitor therapy in HIV-infected men.* AIDS, 2003. **17**(17): p. 2479-86.

424. Cook, R.L., et al., *Analytic strategies to evaluate the association of time-varying exposures to HIV-related outcomes: Alcohol consumption as an example.* Curr HIV Res, 2016. **14**(2): p. 85-92.
425. Ahdieh, L., et al., *Selection by indication of potent antiretroviral therapy use in a large cohort of women infected with human immunodeficiency virus.* Am J Epidemiol, 2000. **152**(10): p. 923-33.
426. Hill, A.B., *The Environment and Disease: Association or Causation?* Proc R Soc Med, 1965. **58**: p. 295-300.
427. Rasmussen, L.D., et al., *Myocardial infarction among Danish HIV-infected individuals: population-attributable fractions associated with smoking.* Clin Infect Dis, 2015. **60**(9): p. 1415-23.
428. Pedersen, K.K., et al., *Microbial translocation in HIV infection is associated with dyslipidemia, insulin resistance, and risk of myocardial infarction.* J Acquir Immune Defic Syndr, 2013. **64**(5): p. 425-33.
429. Armah, K.A., et al., *Prehypertension, hypertension, and the risk of acute myocardial infarction in HIV-infected and -uninfected veterans.* Clin Infect Dis, 2014. **58**(1): p. 121-9.
430. Mason, J.C. and P. Libby, *Cardiovascular disease in patients with chronic inflammation: mechanisms underlying premature cardiovascular events in rheumatologic conditions.* Eur Heart J, 2015. **36**(8): p. 482-9c.
431. Hema, M.N., et al., *Low CD4/CD8 Ratio Is Associated with Non AIDS-Defining Cancers in Patients on Antiretroviral Therapy: ANRS CO8 (Aproco/Copilote) Prospective Cohort Study.* PLoS One, 2016. **11**(8): p. e0161594.
432. European AIDS Clinical Society, *European Guidelines for the Clinical Management and Treatment of HIV Infected Adults: Version 2*. 2005; Available from: http://www.eacsociety.org/files/2005_eacsguidelines-v-english_oct_dublin_1.pdf.
433. Tien, P.C., et al., *Antiretroviral therapy exposure and incidence of diabetes mellitus in the Women's Interagency HIV Study.* AIDS, 2007. **21**(13): p. 1739-45.
434. Tebas, P., *Insulin resistance and diabetes mellitus associated with antiretroviral use in HIV-infected patients: pathogenesis, prevention, and treatment options.* J Acquir Immune Defic Syndr, 2008. **49 Suppl 2**: p. S86-92.
435. Lichtenstein, K.A., et al., *Statin Use Is Associated With Incident Diabetes Mellitus Among Patients in the HIV Outpatient Study.* J Acquir Immune Defic Syndr, 2015. **69**(3): p. 306-11.
436. Schunemann, H.J., et al., *GRADE guidelines: 18. How ROBINS-I and other tools to assess risk of bias in nonrandomized studies should be used to rate the certainty of a body of evidence.* J Clin Epidemiol, 2018.
437. Guyatt, G., et al., *GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables.* J Clin Epidemiol, 2011. **64**(4): p. 383-94.