The integration of independent national HIV surveillance datasets and application of statistical methods to enhance their public health utility

THESIS

presented for the

DEGREE

of

DOCTOR OF PHILOSOPHY

in the faculty of medicine

(Field of study – Epidemiology)

TIMOTHY RICHARD CHADBORN

Research Department of Infection and Population Health,

UCL Medical School

University of London

and

HIV and STI Department,

Centre for Infections,

Health Protection Agency

July 2012

Declaration

I, Timothy Richard Chadborn, confirm that the work presented in this thesis is my own. Where information is derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

The aims of this thesis were two-fold. Firstly, to develop a robust method to create a fully integrated, national surveillance dataset for human immunodeficiency virus (HIV) infections by linking records from three independent, national HIV case reporting systems. Secondly, to apply statistical techniques, more commonly used in cohort study research, to the integrated dataset to yield more of the potential from the constituent information and increase the public health utility of the data. This demonstrated that an integrated dataset can reduce missing information in each surveillance system and improve information use by combining different data that was previously available only in separate databases.

Using the integrated dataset, I achieved the following: accounted for missing information; described the frequency of CD4 count measurements and associated factors; determined characteristics associated with late diagnosis and consequent mortality; estimated the national incidence of acquired immunodeficiency syndrome (AIDS) and death and the influencing factors; assessed information about the date of starting therapy for each individual; and assessed factors associated with immune defence recovery after the start of treatment. These analyses will be/have been integrated into national surveillance processes as appropriate and used to publish academic papers. Lessons have been learnt for surveillance methodology with regards sharing information and ensuring that data are representative of the whole population.

The large size, wide coverage and prospective nature of the integrated dataset mean that national (and local) policy decisions can be based on information that reflects the national picture rather than unrepresentative and time-bound studies. The dataset also has the power to unpick differences within small population groups. For example, evidence about late diagnoses and mortality has been used to promote the need for earlier HIV diagnosis and is updated annually and used to target local needs and to monitor improvements.

Acknowledgements

I would like to thank my supervisors, Professor Caroline Sabin and Dr Barry Evans who patiently guided and advised me throughout my PhD. Their tenacity to stick with me has been invaluable. They have been extremely dedicated to read numerous versions of each chapter. I have learnt so much from them.

Thanks to all my colleagues and collaborators during my work at the Health Protection Agency. Their input to my scientific development has nurtured my epidemiological skills. In particular, Brian Rice, Katy Harker, Jorgen Engmann, Bela Vatsa, Tom Hartney, Cuong Chau, Victoria Bryan and Ana Da Costa who worked in the SOPHID and CD4 Surveillance teams with me, as well as all of the HARS team. Together we developed the datasets which made this thesis possible. Special thanks to Valerie Delpech and Noel Gill for their direction and support. Thanks also to the great support from the HPA for funding my tuition.

Thanks to all the NHS healthcare workers who collect and report surveillance data – my colleagues and I very much appreciate all of their efforts.

I am also indebted to the thousands of HIV-infected individuals who allow their clinical information to be used for surveillance purposes.

My family and friends have provided so much support through the years and had the sensitivity to stop asking when I was going to be finished! But most of all I have to thank Helen, my wife, who has had unbelievable patience with me and spent many hours alone while I studied. She has never known me without my PhD studies but still agreed to marry me and move to Botswana while I whittled away. She will now get me back – and just in time for our first baby!

Chapters

Contents

Tables and Figures

Chapter 3:

Chapter 5:

Abbreviations

AIDS	Acquired Immunodeficiency Syndrome
AOR	Adjusted odds ratio
ART	Antiretroviral therapy
ApproxART	Approximate date of starting ART
CD4	Cluster of differentiation
CD4 Surveillance	CD4 Surveillance Scheme
CfI	Centre for Infections
CHR	Clinician HIV report form
CI	Confidence interval
DOB	Date of birth
E,W&NI	England, Wales and Northern Ireland
ELISA	Enzyme-linked immunosorbent assay
EstART	Estimated date of starting ART
Eth	Ethnicity
FUP	Follow-up
GUM	Genitourinary medicine
HAART	Highly active antiretroviral therapy
HARS	HIV and AIDS Reports database
HIV	Human Immunodeficiency Virus
HPA	Health Protection Agency
ID number	Identification number
IDU	Injecting drug user/s
IQR	Inter-quartile range
IRR	Incidence rate ratio
LTFU	Loss to follow-up
MSM	Men who have had sex with men
NHS	National Health Service
NNRTI	Non-nucleoside reverse transcriptase inhibitor
NRTI	Nucleoside/nucleotide reverse transcriptase inhibitor
ONS	Office of National Statistics
PI	Protease inhibitor
PYFU	Person years of follow-up
Sdex	Soundex code of surname
SOPHID	Survey of Prevalent HIV Infections Diagnosed
SSA	Sub-Saharan Africa
StartART	Reported date of starting ART
STI	Sexually transmitted infection
TB	Tuberculosis
UK	United Kingdom

Chapter 1. Introduction

1.1 Summary information about surveillance data

Surveillance data are usually less detailed than research data but are collected routinely over time. Such data are therefore useful for describing the spread of infection, planning public health responses and allocating resources. Surveillance data for Human Immunodeficiency Virus (HIV) infections are very comprehensive in the United Kingdom (UK) and include individual-level data (no names), reported voluntarily, for most people diagnosed with HIV infection.

Surveillance systems to provide epidemiological information about HIV infections in the UK have been in place since September 1982. These were initiated following the first case of Acquired Immunodeficiency Syndrome (AIDS) in the UK, diagnosed in a homosexual man and reported in December 1981[1]. Surveillance now consists of several complementary systems that provide information about different aspects of the HIV epidemic in the UK. The complexity of certain aspects of HIV infection, including the long term nature of the infection, diagnosis at different stages of clinical progression, the high cost of treatment and care, and the stigma associated with infection, all demand that complex, complementary and confidential surveillance systems are in place to understand the epidemiology. Many of these systems involve assimilation of data that are routinely collected for clinical purposes on a long-term, national basis to provide timely public health information. However, the surveillance systems are separate because the primary sources of routine data are different; clinics, virology laboratories and immunology/haematology laboratories. HIV

surveillance systems in England, Wales and Northern Ireland (E,W&NI) are co-ordinated by the Health Protection Agency's Centre for Infections (HPA, CfI) and are integrated with those of Scotland and the National Study of HIV in Pregnancy and Childhood. HIV and AIDS surveillance data are used to help target and inform public health strategy, to help allocate resources for treatment and care, and to describe the nature of the epidemic.

HIV surveillance systems in E,W&NI include the following: reporting of new HIV diagnoses, new AIDS diagnoses and deaths; annual cross-sectional surveys of all HIV-infected individuals attending for HIV-related care; reporting of CD4 cell counts of HIV-infected individuals; pregnancies in HIV-infected women and babies born to HIV-infected women (managed by the Institute of Child Health); testing samples to identify likely recent infection (still in initial phase); collation of laboratory viral sequence data from suspected resistance cases (managed by the Medical Research Council); and the Unlinked Anonymous laboratory testing of blood sample residues. The Unlinked Anonymous programme is unique because it provides data that are used to gain understanding of undiagnosed infection (and hence estimates of the undiagnosed and overall population prevalence). Such information is irreversibly unlinked from patient identifiers before HIV testing and therefore not used for direct patient management[2].

Monitoring reports on HIV-infected individuals is fundamental for the determination of the number of newly diagnosed HIV infections and newly diagnosed AIDS patients, the number of deaths among HIV-infected individuals, the number of diagnosed HIV-infected individuals seen for care during a

calendar year, and the level of immunosuppression of these individuals. Surveillance data from E,W&NI that provide this information are available from the following databases:

- HIV and AIDS Reports database (HARS database), which collates reports from virology laboratories of all antibody positive individuals diagnosed with HIV for the first time at that laboratory (lab reports), reports from physicians of HIV-infected individuals newly diagnosed (on clinician report forms (CHRs) since January 2000), reports from physicians of new AIDS diagnoses (AIDS reports that have been adapted into CHRs since January 2000), and reports of deaths among individuals living with HIV from both physicians and the Office of National Statistics (ONS);
- CD4 Surveillance Scheme database (CD4 database) that collates reports from haematology and immunology laboratories of longitudinal CD4 cell counts for HIV-infected adults (men and women aged 15 or more);
- Database of the Survey of Prevalent HIV Infections Diagnosed (SOPHID database), which collates reports of attendees of facilities providing statutory HIV-related treatment and care.

1.2 Confidentiality in sexual health care and HIV surveillance

The National Health Service (NHS) of the UK is a tax funded healthcare system with universal coverage, which is generally free at the point of care. HIV treatment and care is provided by open access and confidential (anonymous if requested) genitourinary medicine (GUM) clinics or specialist HIV clinics because sexual health and HIV remain stigmatised[3].

Due to stigma, HIV surveillance systems in the UK are not based on statutory notifications (like many other infections) but voluntary and 'pseudonymised' HIV surveillance is permitted (Appendix A.1). Although reports pertain to an individual, the data received by the HPA do not contain patient names but a code based on the surname (but not unique to a given name) and date of birth and are therefore referred to as 'pseudonymised' (i.e. partially anonymised). To ensure confidentiality, all HIV-related data are stored on restricted and secure databases at the HPA, with strict adherence to the Data Protection Act and Caldicott Guidelines[4]. No patient-level data are ever released by the HPA, except back to clinics who have reported that information, and aggregate data are only published after ensuring that deductive disclosure cannot occur.

1.3 Background to HIV infections

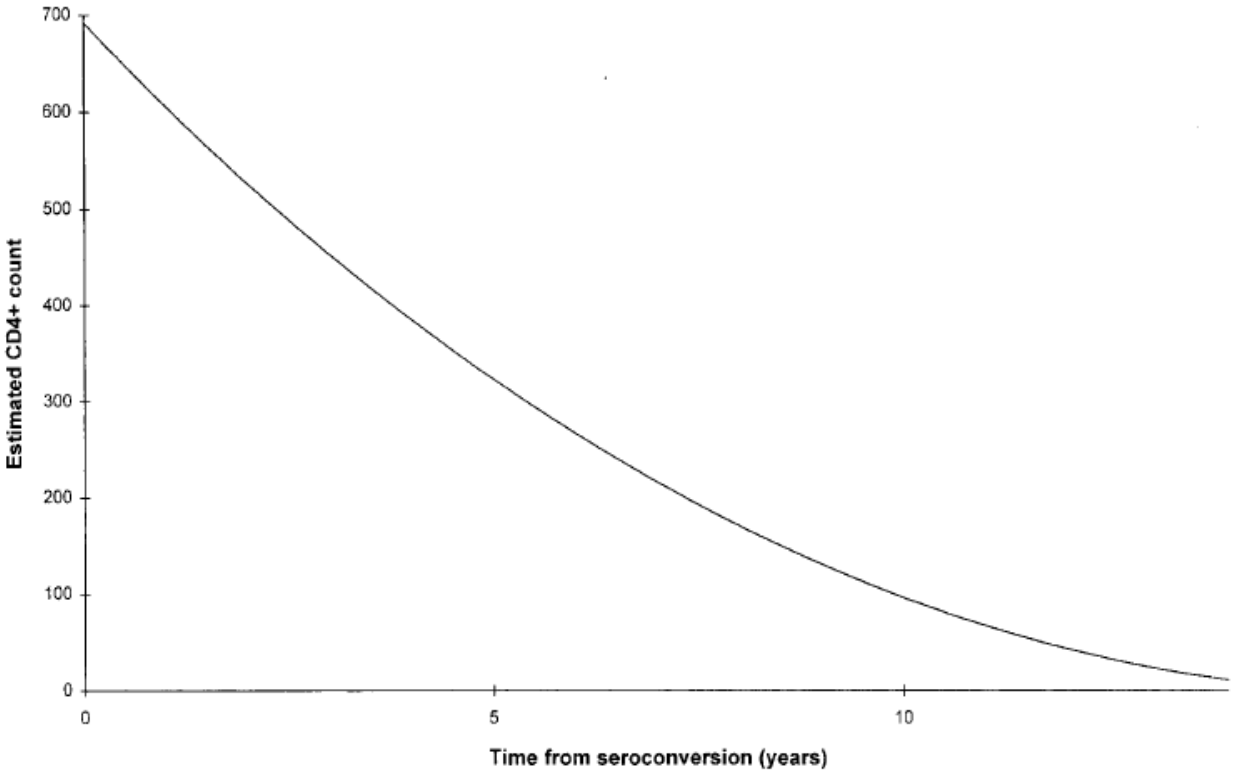
1.3.1 Overview of HIV natural history and transmission

HIV was first identified in 1983 as the causative agent of AIDS[5-10] two years after AIDS was first recognised among homosexual men[11]. HIV is a chronic infection that appears to be terminal in all cases in the absence of treatment. Acute infection has been identified as a short phase of unspecific symptoms (fever, rash and sore throat) in 40-90% of adult cases but these symptoms are not usually recognised as primary HIV infection[12]. There often follows an extended period (sometimes many years) without symptoms before the development of serious clinical events and/or death. During this period the infected individual becomes progressively more immunodeficient until opportunistic infections cause illness and eventual death. AIDS represents the

late clinical stage of HIV infection and is clinically defined as a spectrum of opportunistic infections, cancers, neurological complications and other disorders (Appendix A.2)[13;14]. Not all infected individuals who die with HIV infection have developed AIDS, although the vast majority did so prior to the availability of effective treatment.

The exact mechanism by which HIV infection impairs the human immune system remains unclear and it is likely that a combination of proposed mechanisms is involved. However, between 1981 and 1987 the CD4 molecule on the surface of T helper lymphocytes was identified as the receptor for the HIV virus[15-17], the immunodeficiency associated with AIDS was shown to be due to a progressive loss of these CD4 cells due to HIV infection[18-23], and CD4 cell counts were shown to be predictive for the development of AIDS[24-28]. CD4 cell counts (absolute T-helper lymphocyte count as a concentration per cubic millimetre of blood) are therefore most frequently used as markers of immunodeficiency and the risk of progression. Declining CD4 counts as a function of time are illustrated in Figure 1.1, which represents the natural history of HIV infection. The literature suggests that this trend is best described by a curve, which is linear in the square root or logarithm of the CD4 cell count[29-35].

Figure 1.1. Natural history of infection shown by declining CD4 counts*



The function of the CD4 cells and the ability of the thymus to generate CD4 cells in the presence of HIV infection may be important determinants of the rate of progression of infection[36;37]. Decreasing thymic function with age has been recognised as a major cause of increased progression with age[38-41]. However, CD4 cell function is infrequently studied in the routine medical care of HIV-infected individuals.

* Reproduced from Cozzi Lepri et al. Journal of Infectious Diseases 1997;175:775–80 [29]

1.3.2 CD4 cell counts prior to HIV infection and measurement error

CD4 cell counts tend to be stable for each adult without HIV infection over long periods of time[42]. Yet, there is a wide range of CD4 cell counts among uninfected adults (studies indicate medians between 591 and 1116 cells/mm^3 with large inter-quartile ranges [IQR])[42-47]. Women have higher CD4 cell counts than men[43;44;48-53] but there is no consensus on whether these differences are maintained among HIV-infected individuals or whether they are associated with an effect on the rate of progression to AIDS or death among untreated individuals[49;54-56] (the similar progression rates for men and women suggest that the difference becomes less pronounced over time since infection[49]). There is also evidence that CD4 counts may be up to 30% lower during pregnancy but that the trend is u-shaped and CD4 counts recover to previous levels after delivery[50;57]. There is some evidence for differences in the CD4 cell counts of uninfected individuals by geography/race/ethnicity. Most comparisons have been between summary results of studies and largely indicated lower CD4 counts among Asians, Africans, Middle-Eastern and Chinese than among Caucasians[42;53;58-64] although similar studies have contradicted these findings[65-69]. Studies that directly compared racial/ethnic groups either did not control for possible confounding factors[70-75], did not find any differences by race/ethnicity[76], or found no differences by race/ethnicity after controlling for other factors[77]. It is important to note that even if there are differences between races/ethnicities in CD4 cell counts prior to HIV infection, these do not have any significant clinical effect on the rate of progression to AIDS or death

because of faster CD4 decline for those with higher CD4 counts [63;78-83]. Furthermore, CD4 cell counts prior to infection are rarely known for HIV-diagnosed individuals. Therefore, CD4 cell counts prior to infection are not used at the population or individual-level to vary the immunologic threshold for initiation of antiretroviral therapy.

Other factors are associated with CD4 cell counts among adults without HIV infection but only account for a small proportion of the observed variability. There is a diurnal variation such that CD4 cell counts increase during the day[48], and pregnancy, use of the contraceptive pill, smoking, illness (including tuberculosis and viral infections such as influenza), and trauma have all been associated with temporarily diminished CD4 cell counts[42-46].

There is likely to be significant measurement error in the determination of CD4 cell counts as there is considerable variability both within an individual and between laboratories[48]. This variation is likely to have decreased over time as more reproducible laboratory techniques have developed.

1.3.3 Transmission, viral loads and viral types/subtypes

Susceptibility to HIV infection appears to be universal although host factors, co-infections and medical interventions can affect the risk. Documented routes of infection are sexual exposure, exposure to blood or tissues and breast feeding with infectivity correlated with the amount of virus (viral load) in the bodily fluid.

Often in the second week after infection, a peak is seen in the viral load followed by the development of a specific immune response[84;85]. There is

continuing debate as to whether the viral load tends to fluctuate around a stable 'set point' for some years or whether viral load levels progressively increase over the subsequent years of infection[86-90]. One theory suggests that the 'set-point' varies significantly between individuals reflecting different steady-state levels of viral replication. Subsequent increases in viral load, related to a failure of the immune system to keep viral replication in check, often then predict the development of symptomatic disease. The contrasting theory suggests that viral loads rise progressively, and with it, the risk of AIDS and death[88;91-93].

There are three main types of HIV virus, HIV-0, HIV-1 and HIV-2, which are serologically and geographically relatively distinct with numerous subtypes. HIV-1 is most commonly found in the UK whereas HIV-0 and HIV-2 remain mainly confined to West Africa. By the end of 2007, there were 121 diagnoses of HIV-2 infection in individuals without HIV-1 infection reported in the UK and a further 22 diagnoses of HIV-1 and HIV-2 dual infections. HIV-1 is more pathogenic than HIV-0 or HIV-2 with a faster rate of CD4 cell count decline, a faster rate of disease development and shorter survival[94-96]. There is also evidence that infection with HIV-1 subtype D or some recombinant subtypes is associated with a higher risk of death and faster CD4 cell count decline than subtype A infection[80;97-103]. However, HIV types will not be considered in this thesis because HIV testing strategies do not always seek to distinguish types and subtypes and because surveillance lacks good data in this area.

1.3.4 Detection of HIV/HIV tests/diagnosis

Detection of the HIV virus has been widely carried out since 1984 using the enzyme-linked immunosorbent assay (ELISA), which detects antibodies specific to HIV in venepuncture serology samples. '4th generation' HIV tests, which include antigen detection to increase sensitivity are now the standard of care. Point-of-care testing kits, which use blood from fingerpricks or mouth swabs and provide results within minutes, can facilitate HIV testing where laboratory testing is not feasible or when results are required promptly. However, these currently do not detect antigen and have lower specificity. Independent assays should be used to confirm all positive HIV test results[104;105].

In the UK, HIV testing has been primarily available at sexual health, HIV and antenatal clinics. Prior to 1999, HIV testing was largely a result of individuals requesting an HIV test. Subsequently, HIV testing increased with the recommendations that the following groups should be offered a test: all pregnant women (1999); all GUM attendees (2001); all individuals with symptoms indicative of HIV infection attending healthcare settings (2001); all individuals with a high risk of infection* (2008); all adults registering in primary care and all general medical admissions where diagnosed HIV prevalence estimates in the local population exceed 2 in 1000 (2008)[105-111].

* all men and women reporting a history of IDU, all individuals known to be from a high prevalence country or who report sexual contact with individuals from areas of high HIV prevalence, and all men who have disclosed sexual contact with other men

1.3.5 Key risk groups for HIV infection

Men who have had sex with men (MSM), injecting drug users (IDU) and individuals from areas of high HIV prevalence, particularly sub-Saharan Africa (SSA), are preferentially offered HIV testing because these groups are disproportionally affected by HIV infection in the UK. Estimated overall HIV prevalence among individuals aged 15 to 44 years in England, Wales and Scotland in 2006 was 5.36% among MSM, 0.43% among IDU, 2.86% among women born in SSA, 1.57% among men born in SSA, 0.04% among women born elsewhere and 0.03% among men born elsewhere (unpublished HPA data using Multi-Parameter Evidence Synthesis methodology[112-114]). Additionally, many individuals were unaware of their infection (Table 1.1)[115].

Table 1.1. Estimated[*] number of adults (15-59 years) with HIV in the UK in 2006

Exposure category	Number diagnosed[†]	Number undiagnosed	Total
MSM	**20,900** (20,300, 21,500)	**9,200** (6,800, 12,800)	**30,100** (27,600, 33,700)
IDU	**1,600** (1,400, 1,900)	**1,100** (700, 1,600)	**2,700** (2,300, 3,300)
Heterosexuals	**25,300** (24,500, 26,100)	**11,100** (7,600, 16,200)	**36,400** (32,900, 41,400)
Men	**9,100** (8,800, 9,400)	**5,600** (3,600, 9,300)	**14,700** (12,700, 18,400)
African born	*5,700 (5,500, 5,900)*	*3,200 (1,600, 6,400)*	*8,900 (7,300, 12,100)*
Non-African born	*3,400 (3,200, 3,500)*	*2,400 (1,600, 3,900)*	*5,800 (5,000, 7,300)*
Women	**16,200** (15,700, 16,800)	**5,400** (3,700, 7,500)	**21,600** (20,000, 23,700)
African born	*12,100 (11,700, 12,500)*	*3,500 (2,000, 5,300)*	*15,600 (14,100, 17,400)*
Non-African born	*4,200 (4,000, 4,300)*	*1,800 (1,200, 3,000)*	*6,000 (5,300, 7,200)*
Grand total	**47,800** (46,500, 49,100)	**21,600** (17,000, 27,800)	**69,400** (64,800, 75,500)

[*] Numbers of prevalent infections (rounded to the nearest 100) with credible ranges were estimated using Multi-parameter Evidence Synthesis, in an extension of the method described in Goubar A et al [112]

[†] Numbers diagnosed exclude individuals aged 15-59 living with HIV in 2006 infected through blood or blood products or tissue (466) or through mother to infant transmission (233). Numbers diagnosed include individuals with unknown exposure, allocated according to the distribution of those with known exposure.

1.3.6 Treatment of HIV-infected individuals

Daily therapy with a combination of antiretroviral drugs can reduce the viral load below the level of quantification, increase CD4 cell counts, and reduce the risk of opportunistic infections and the development of AIDS, which prolongs life[116]. However, insufficient adherence allows viral mutations to develop that can promote resistance to antiretroviral drugs and loss of the benefits of treatment[117]. It is often difficult for patients to adhere sufficiently even if they do not suffer from any side effects of therapy, but these are common and can cause significant morbidity, discouraging patients from starting or continuing treatment. Therefore, treatment of an HIV-infected individual must balance the advantages of attaining viral suppression with the risks of drug toxicities and viral resistance. Even assuming no resistance or toxicities, it would take an estimated 73 years to eliminate HIV from the body with currently available therapies because of a latent reservoir of HIV in resting memory CD4 cells[118-120]. Thus, it is unlikely that HIV clearance will be achieved and therapy is expected to be lifelong. Yet, the majority of patients can tolerate antiretroviral therapy (ART) well for years and treatment can be effective even after detection of resistance[121;122].

There are currently five classes of antiretroviral HIV drugs approved in the UK: nucleoside and nucleotide reverse transcriptase inhibitors (NRTIs), non-nucleoside reverse transcriptase inhibitors (NNRTIs), protease inhibitors (PIs), fusion (or entry) inhibitors and an integrase inhibitor. Most effective regimens

consist of a combination of drugs, referred to as 'highly active antiretroviral therapy' (HAART), and often made up of two NRTIs and either an NNRTI or PI.

There is much debate about the optimum time to start therapy due to the need for a balance between viral suppression and the risks of drug toxicities, viral resistance and long-term damage to the immune system. Treatment guidelines recommend that HAART be initiated fairly late in the course of infection, primarily based on CD4 cell counts, but before the patient has become at markedly increased risk of AIDS or death[123]. Cohort studies have shown that the short-term risk of AIDS or death is higher for individuals starting treatment with CD4 cell counts below 200 cells/mm^3 than for those with counts over 200 cells/mm^3 and that there is little further decline in the risk as counts increase to over 350 cells/mm^3 or over 500 cells/mm^3 [124;125]. Therefore, British guidelines between 2001 and 2006 recommended that individuals start treatment before CD4 counts fell below 200 cells/mm^3 but that treatment was deferred for most until counts were below 350 cells/mm^3 (also that treatment be initiated immediately for individuals with AIDS defining illnesses and considered for individuals with high viral loads or rapid loss of CD4 cells) (Table 1.2). However, the recommended threshold changed over time as new evidence informed the point at which the benefits of treatment were likely to outweigh its risks[126;127]. Randomised controlled trials to determine the optimal time to initiate therapy have not yet been completed. In 2008 and 2012 (after the period for which this thesis analyses data), these guidelines were updated again. In 2008, the guidelines were updated to recommend the initiation of treatment for all individuals with CD4 count below 350 cells/mm^3 and also for those with CD4

count between 350 and 500 cells/mm^3 with higher risk of clinical events[*] [123]. These guidelines remained very similar in 2012 except that CD4 counts close to the threshold of 350 cells/mm^3 were included in the recommendation for initiation of treatment and the limit of 500 cells/mm^3 was removed for individuals indicated treatment for hepatitis B and for patients wishing to reduce the risk of transmission[128].

Table 1.2. Recommended CD4 thresholds for initiating ART.

Year guidelines published	CD4 threshold (cells/mm^3)	Recommendation (*paraphrased*)
1997[129]	<300	Consider at all levels. No evidence to indicate the optimum time to start therapy.
1998[130]	>350	Should begin before irreversible damage has occurred to the immune system such as at a CD4 count above 350 cells/mm^3.
2000[131]	<350	Treat when CD4 count is less than 350 cells/mm^3. Consider when CD4 count is 350-500 cells/mm^3.
2001[132]	200-350	Start treatment within 200-350 cells/mm^3 range.
2003[133]	201-350	Most individuals should start therapy when the CD4 count is in the range 201-350 cells/mm^3.
2005[134]	200-350	Recommended that the majority of people initiate therapy with CD4 counts of 200 to 350 cells/mm^3
2006[135]	200-350	Treatment should start before CD4 cell count has fallen to below 200 cells/mm^3. Majority of patients with CD4 cell counts greater than 350 cells/mm^3 should continue to defer treatment.
2008[123]	≤350	Treat as soon as possible when CD4 count is less than 350 cells/mm^3. Consider in specific situations when CD4 count is 350-500 cells/mm^3.
2012[128]	≤350	Treat when CD4 count is less than 350 cells/mm^3 or close to this threshold. Consider in specific situations when CD4 count is >350 cells/mm^3.

[*] any HIV-related co-morbidity, hepatitis B or C co-infection, low CD4 percentage, established CVD or a very high risk of cardiovascular events, or if a patient wished to reduce the risk of transmission.

Treatment, although expensive, is free to patients in the UK (although only free for undocumented migrants in emergency situations). It is provided from open-access, voluntary and confidential HIV-specific clinics in most areas and often integrated with genitourinary medicine clinics, which provide sexual health services[136].

1.4 Summary of aims of this thesis

The aims of this thesis were firstly, to develop a robust method to create a fully integrated, national surveillance database for HIV infections by developing methods to link records from three independent, national HIV case reporting systems. And secondly, to apply statistical techniques, more commonly used in cohort study research, to the enhanced database to enable the constituent information to be analysed to its full potential, increasing the public health utility of the data.

In chapter 2, I describe the methodology of the three national HIV case reporting systems, compare these systems to those of other developed countries, and describe the epidemiology of diagnosed HIV infections in E,W&NI using the national surveillance data. In chapter 3, I describe how I linked records from the three surveillance systems by comparing patient-specific information (such as year of birth) between datasets using a hierarchical algorithm. This allowed for some clerical errors in the data and only retained links between the datasets that were reliable enough for use in analyses. Chapter 4 describes how an integrated dataset containing a single

patient record consisting of a coherent sequence of events was created from the linked records by triangulation and data validation. Chapter 5 shows how I investigated potential sources of bias that could affect subsequent analyses. These were inclusion bias and bias due to differential CD4 cell count monitoring, loss to follow-up (LTFU) and informative censoring. The aim of chapter 6 was to investigate the frequency of 'late diagnosis' of HIV infection and its consequent effect on mortality within a year of HIV diagnosis by analysing CD4 cell counts at the time of HIV diagnosis (as a marker of diagnosis after the recommended time for treatment initiation). In chapter 7, I use longitudinal CD4 cell counts for each individual to allocate follow-up time and to estimate the national incidence of AIDS and death and the influencing factors, considering specifically the first six months after HIV diagnosis. In chapter 8, I investigated a number of algorithms to estimate the time of starting ART based on changes in the CD4 cell count, with the aim of supplementing surveillance data for individuals for whom this data had not been reported. Chapter 9 describes the rate of change of CD4 cell counts after the initiation of ART and the determination of factors associated with these rates. Chapter 10 provides a brief summary of the thesis with lessons learnt and further work indicated.

The work included in this thesis was conducted over a number of years, over which time the epidemiology of HIV in the UK changed markedly and new data fields were added to the dataset. Thus, all data and results were updated in 2008.

Chapter 2. HPA case report systems and HIV epidemiology in England, Wales and Northern Ireland

2.1 Introduction

This chapter describes the data and case report systems used for monitoring HIV infections among adults in England, Wales and Northern Ireland (E,W&NI). The systems are the HIV and AIDS Reporting System (HARS), the Survey of Prevalent HIV Infections Diagnosed (SOPHID), and the CD4 Surveillance Scheme (CD4 Surveillance), which receive data from clinics, virology laboratories and immunology laboratories.

The second section uses data received by these systems to describe the HIV epidemic. It highlights the limitations of surveillance and the conclusions that can be drawn by making informed assumptions about the data. It shows that MSM have continuously been diagnosed in large numbers but that estimated numbers of MSM living with HIV in the UK have recently been almost matched by black African heterosexuals from sub-Saharan Africa. As well as showing geographical foci of the epidemic, case reports indicate that prevalence has increased markedly in association with increasing numbers of new diagnoses, and decreasing numbers of deaths due to treatment.

Section three summarises HIV case reporting in other countries with similar HIV epidemics to provide context and a wider understanding of HIV surveillance.

2.1.1 Data analysed: HIV infections among adults in E,W&NI to the end of 2007

The data analysed throughout this thesis was derived from reports of HIV infections among adults (aged 15 years or more) from clinics or laboratories in E,W&NI. Transmission of HIV from mother to child was first recognised in 1982[137] and case reports due to this route of infection were originally included in the remit of the HPA but taken on by the Institute for Child Health in the late 1980s. Reports are added to the HPA databases on a regular basis and included in UK surveillance reports. However, missing patient identifiers on reports for children limits the possibilities for data linkage across surveillance systems. Additionally, the natural history of infection among children is quite different than it is among adults.

Case reports from Scotland are made to Health Protection Scotland and aggregate data are included in UK surveillance reports. However, only disaggregate data on new diagnoses in Scotland have been historically incorporated into the HPA database so Scottish data was not included in the data analysed. Analyses were updated in 2008, when data were practically complete for 2007 so events after the end of 2007 were not considered.

2.1.2 Key data used to link records from the same individual

Key data for HIV case report surveillance systems in E,W&NI include the 'pseudonymised' patient identifiers: soundex code, date of birth, firstname initial, area of residence (reported as postcodes [Appendix A.3]), local patient identification (ID) number, and demographic information on probable route of

infection (subsequently referred to as risk group) and ethnicity. These data are fundamental to individual-based HIV surveillance systems and to this thesis because, in combination, they are considered specific enough to link records from the same individual while maintaining confidentiality because no names are collected.

The soundex is a four-character alphanumeric code of the patient's surname (e.g. C316 for Chadborn [Appendix A.4]) based upon phonetic grouping of its consonants, which is not unique to a surname due to redundancy in the coding system (only 6,734 possible soundex codes are available). Therefore, it is not possible to decode a soundex code to a specific surname (to reverse the function) with any certainty. Soundex codes were originally developed with the aim of matching equivalent surnames that had been spelt differently over time or because of clerical errors[138].

Local patient ID numbers (also known as clinic numbers but not to be confused with numbers identifying clinics) are assigned by HIV and sexual health clinics to uniquely and anonymously identify patients locally. HIV surveillance systems use these numbers to identify reports from the same individual originating from the same clinic or hospital. Local patient ID numbers do not have a nationally standardised format although many are alphanumeric codes that include an incremental number, which increases as new patients are added to the clinic cohort, and prefixes and/or suffixes, which may denote the patients' sex, year of first attendance at the clinic and/or the clinic within the hospital (e.g. GUM99M00132).

The frequently long delay between infection and diagnosis means that individuals often do not know exactly when or from whom they acquired infection. To understand the epidemiology of HIV in the UK, reports to the HPA are categorised by their most likely route of infection according to a hierarchy (Appendix A.5) related to the decreasing risk of transmission (assigned according to data on reports). Individuals categorised as having acquired their infection heterosexually are sub-categorised according to the risk group of their partner(s) following the same hierarchy. Those presumed heterosexually infected by a partner presumed also to have acquired infection heterosexually are again sub-categorised. This sub-categorisation is according to whether the exposure occurred abroad or in the UK, and then by world region of exposure (ranked by HIV prevalence). Exhaustive follow-up is attempted for all individuals who cannot be categorised from data initially reported (Sections 2.2.9 and 2.4.3).

The ethnicity of HIV-infected adults was requested on all AIDS reports, on all HIV reports since 2000 (since 1993 for individuals infected abroad) and on all clinician reporting forms (CHRs). Ethnicity was not routinely followed-up if missing but was requested if follow-up of other data was undertaken. It was categorised as follows: i) White, ii) Black African, iii) Black Caribbean, iv) Black – other, v) Indian\Pakistani\Bangladeshi, vi) Other/mixed.

2.2 Reporting of HIV and AIDS diagnoses and deaths

2.2.1 Background

A report to provide detailed information about AIDS in the UK was proposed in August 1982[139] following a case* of AIDS in a homosexual man in December 1981[1]. In September 1982 a voluntary clinical reporting system was initiated by requesting consultant dermatologists and venereologists from E,W&NI to report new diagnoses of AIDS to the HPA's Centre for Infections[141]. In November 1982, microbiologists were asked to include age, sex and sexual orientation on their reports of Kaposi's sarcoma and severe opportunistic infections when the condition appeared to fit the definition of AIDS. Also in 1982, the Office of National Statistics (ONS) began to provide the HPA with details of deaths due to Kaposi's sarcoma or AIDS on a weekly basis. Although the case definition of AIDS has been revised since the early 1980s, these original methods remain essential to the surveillance of HIV and AIDS in the UK today (Figure 2.1).

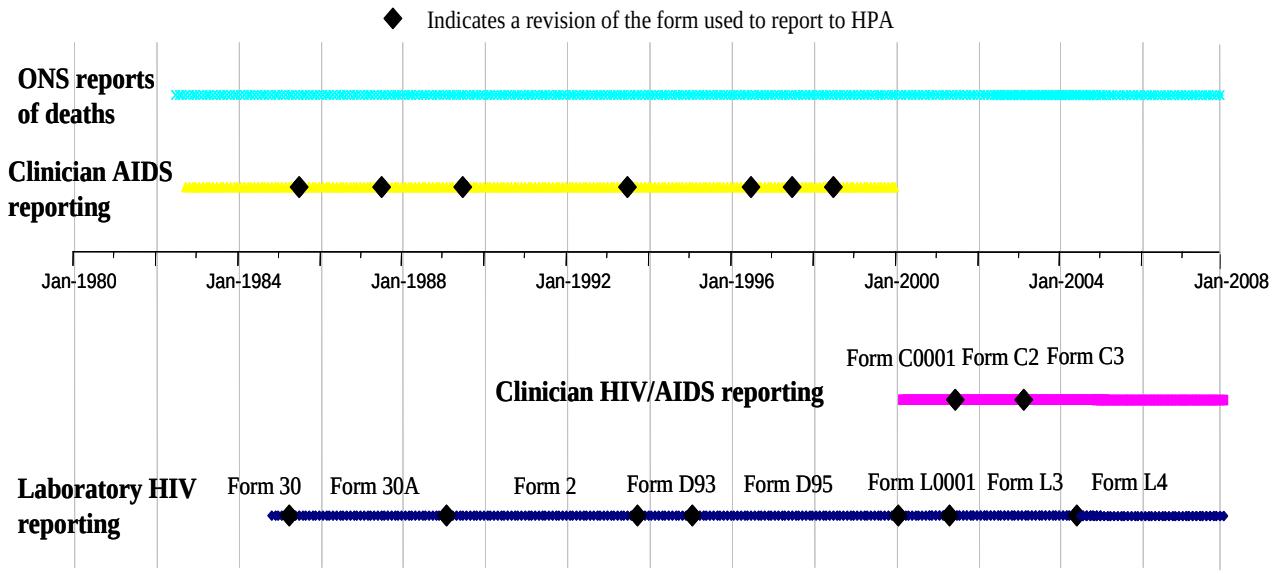
HIV antibody tests became available at the HPA's Virus Reference Laboratory in the UK in late 1984[142] and voluntary HIV serosurveillance was added to national surveillance in March 1985 (Section 2.2.2). Virologists who receive specimens for HIV testing have since been asked to report to the HPA all

* based on a working definition of AIDS among adults developed by the Centers for Disease Control and Prevention, Atlanta, for the purposes of epidemiologic surveillance [140]. Cases were defined "as illness in a person who (1) has either biopsy-proven Kaposi's sarcoma or culture-proven, life-threatening opportunistic infection, (2) is under age 60 and (3) has no history of either immunosuppressive underlying illness or immunosuppressive therapy".

individuals when they are first identified as HIV positive in their laboratory. Detection of HIV also enabled retrospective HIV diagnoses from stored samples such as those from haemophiliacs. Information about these HIV infections was collated in the database of the United Kingdom Haemophilia Centre Doctors' Organisation and epidemiological data on new diagnoses were sent to the HPA.

A timeline of the development of the various surveillance methods for reporting HIV and AIDS diagnoses and deaths is shown in Figure 2.1. The methods and forms used are described in more detail in the following sections.

Figure 2.1. Timeline of the introduction and modification of the various surveillance methods for reporting HIV and AIDS diagnoses and deaths.



The HARS database is managed by a team of epidemiologists and data management staff at the HPA. This usually consists of a two data entry clerks, two graduate scientists and a research nurse led by a Masters-level epidemiologist and overseen by a medical consultant epidemiologist.

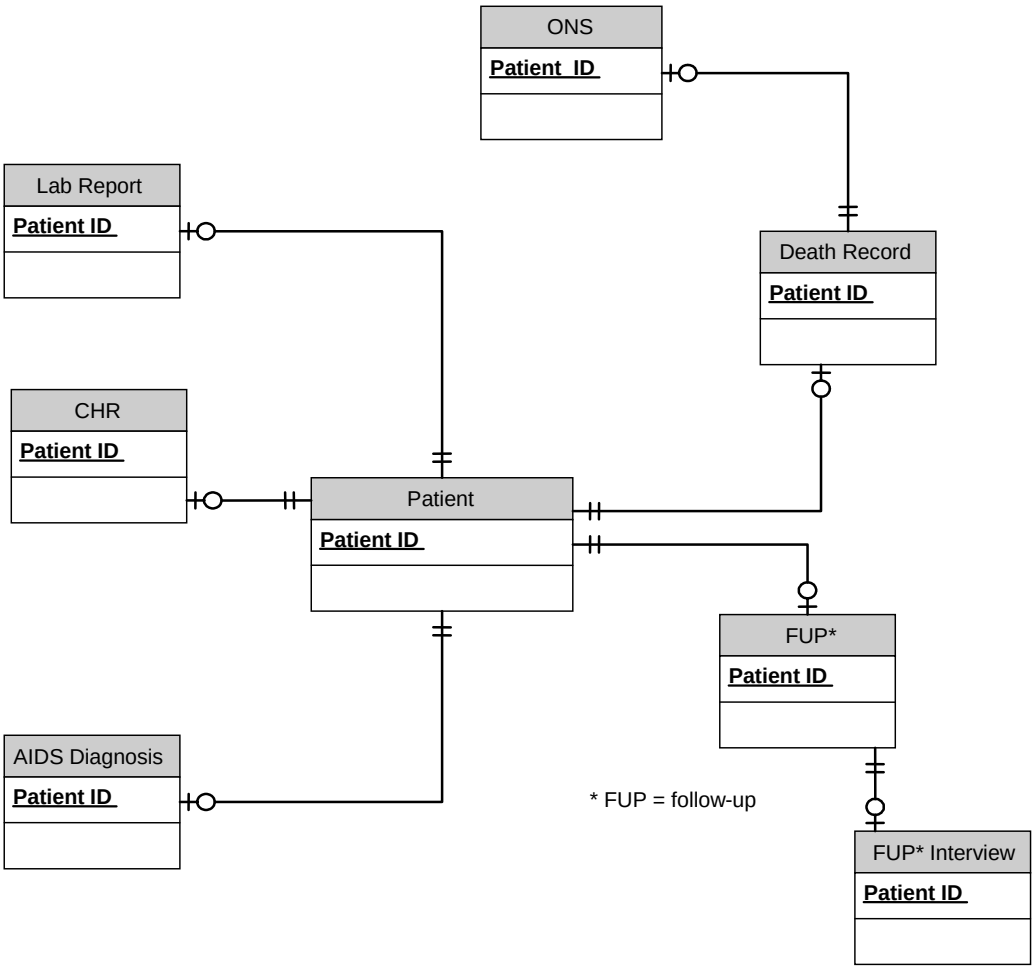
2.2.2 Ascertainment of reports of HIV and AIDS diagnoses

HIV and AIDS reporting have always been voluntary and in strict medical confidence so as not to deter individuals from being tested. Whilst largely complete, an estimated 10-20% of new diagnoses were not reported to the HPA in the first decade of the epidemic[143]. However, this compared favourably with the completeness of surveillance for infections made officially notifiable[144]. Information as to why reports are not sent is not available and the proportion under-reported has not been robustly determined.

2.2.3 The HARS database

Databases have been used to support the surveillance of HIV/AIDS since the establishment of AIDS reporting in 1982 and data have been transferred between databases at the time of each upgrade (every five years since 1985) (Appendix A.6). A separate HIV database was created with the advent of HIV serosurveillance in 1985 but since 1995 the HIV and AIDS databases have been unified. These databases were patient-based with a single integrated and coherent record for each patient linked to reports of HIV diagnoses, AIDS diagnoses and deaths. Integration of the AIDS and HIV databases allowed simple production of national figures of the number of individuals reported as HIV-diagnosed. The structure of the database, showing the main data tables and their relationships at the end of December 2007 is shown in Figure 2.2.

Figure 2.2. Main data tables and their relationships in the database of HIV diagnoses, AIDS diagnoses and deaths for HIV-infected individuals: 2007.



The ‘patient’ table forms the basis of the database with a single integrated and coherent record from each patient containing data from potentially a number of sources. This source information is recorded, as reported, in the supporting tables according to the source and type of form: the ‘lab report’ table stores minimal data about newly detected positive HIV test results that are reported on laboratory HIV reporting forms from virology laboratories; the ‘AIDS diagnosis’ table stores data about new AIDS diagnoses reported on AIDS forms by clinicians; the ‘CHR’ table stores both data on new HIV infections and new AIDS

illnesses diagnosed by clinicians that are both reported on CHRs; the ‘death record’ table stores a single integrated and coherent record from each patient who is known to have died irrespective of the source of that information; the ‘ONS’ table stores records of deaths reported to ONS that have been linked to records of HIV-infected individuals in the ‘patient’ table; the ‘FUP’ (follow-up) table stores records of reports that require follow-up due to incomplete information; the ‘FUP interview’ table stores information from interviews of patients conducted to ascertain that incomplete information. The ‘patient ID’ links all records from the same patient.

On addition of new reports to the database, data are merged if reports are complementary (HIV and AIDS reports, HIV and death reports, AIDS and death reports, or HIV, AIDS and death reports). If the same event is reported more than once then only the report with the earliest date of event is retained in the patient record. However, any different demographic information on the report to be deleted is used to update missing data on the report that is retained. Conflicting information is dealt with individually and on an ad-hoc basis as new report forms are received. If a clear majority of reports concur and differ from the new information then the new information is not used to update the patient record. However, demographic data reported directly from clinicians is assumed to be more dependable than data routed through laboratories. If a clear decision cannot be made then follow-up is undertaken to resolve the inconsistency.

Since 1997, an additional and separate flat table, called ‘XLatest’ has been created in the database at the end of every quarter (biannually since January

2008). This aggregates key elements of all reports received to date into one record per individual and is therefore more comprehensive than the 'patient' table (Appendix A.7). The 'earliest event date' in this table is understood to be the earliest reported date of confirmed HIV-infected status for the individual whether it comes from a laboratory report, a clinician report or a death report (assuming complete reporting of events). For the large majority of records, this reflects the first positive result from voluntary testing. In a few cases this is the date of clinical AIDS diagnosis prior to an HIV test (in particular for those cases diagnosed before the availability of antibody tests) and in even fewer, a date of death where post-mortem investigation reveals evidence of HIV-infection. Related information, such as the 'earliest organisation' that reported the event and the 'age at earliest event' are extracted from the relevant forms, which are recorded in the 'earliest event code'. The first reported date of birth and soundex code are assigned to the patient record, unless subsequently reported or found through follow-up to be incorrect.

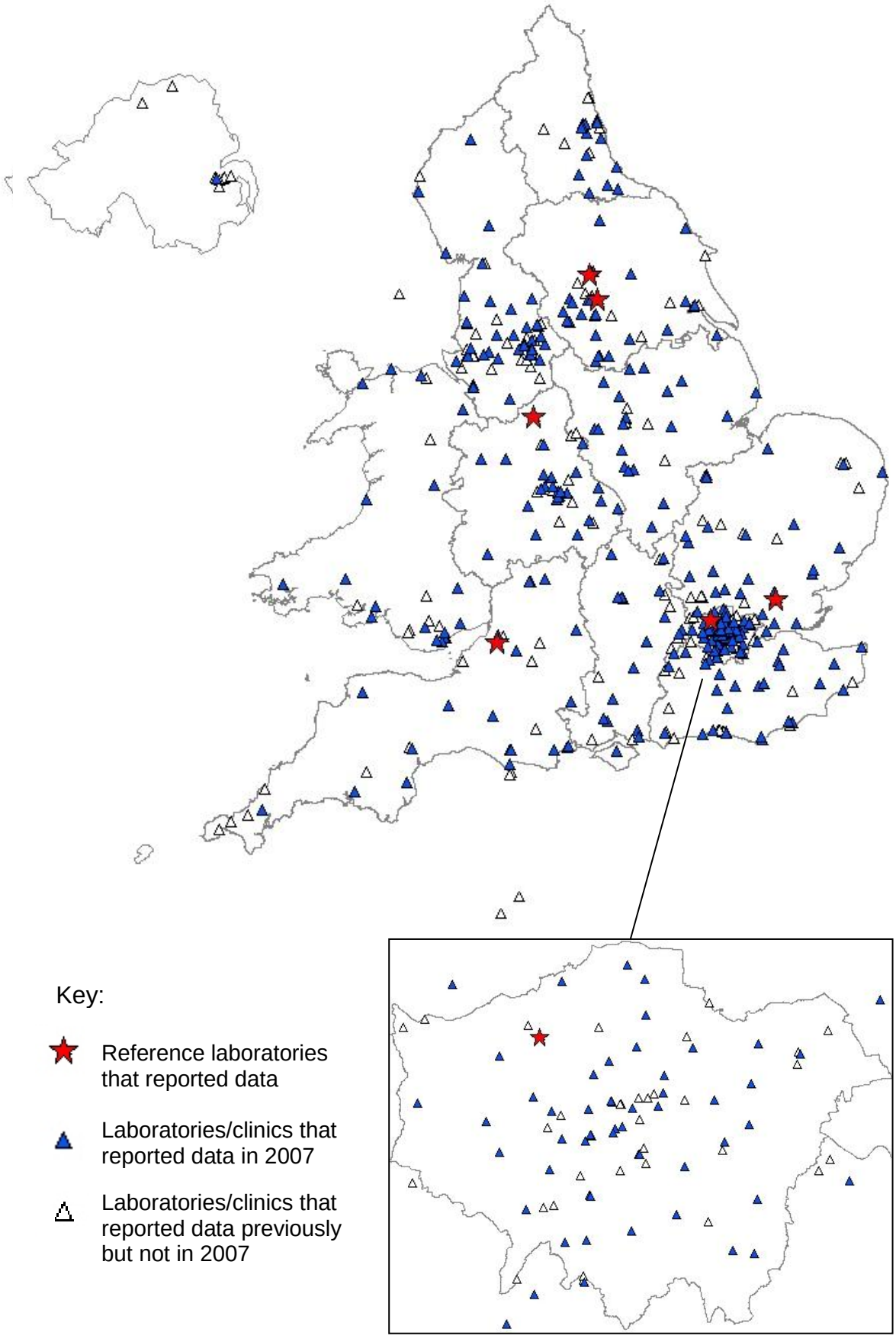
2.2.4 Laboratory reporting of HIV and AIDS diagnoses

In 1984, microbiologists began to report newly diagnosed HIV infections individually on pre-existing forms for the reporting of other infections (Appendix Table B.1 and Appendix Forms B.1a – B.1i). After HIV serosurveillance was initiated in March 1985, an HIV-specific form was used until 1989 that allowed the reporting of up to 9 cases per sheet and was accompanied by a specific guide for completion. However, it was not until September 1993 (after reincorporation into the standard laboratory reporting form for individual cases)

that the first dedicated HIV form with accompanying guidance sheet was introduced. This was used with minor modifications until January 2000 when clinician reporting of newly diagnosed HIV infections was introduced. As many diagnoses would subsequently be reported by clinicians it was therefore no longer justifiable to request information that was not readily available in the laboratory and was expected to be duplicated on the clinician HIV form (CHR). However, laboratories were asked to continue reporting newly diagnosed HIV infections to indicate diagnoses that were not reported by clinicians and to provide preliminary information in the case of delayed clinician reporting. Multiple records of an HIV diagnosis for an individual are expected because all laboratories are asked to report the first time that they diagnose an individual even if they know that the patient was previously diagnosed elsewhere. This is to maximise the completeness of surveillance and to save reporters from enquiring whether patients have been diagnosed elsewhere. However, if patients are known to have been diagnosed elsewhere, the form requests the date and place of earliest diagnosis to help identify and link related reports.

A total of 417 laboratories (serving 621 clinics) in E,W&NI had been involved in reporting newly diagnosed HIV infections among adults to the HPA to the end of June 2008 (Map 2.1). By the end of June 2008, 206 laboratories (serving 311 clinics) had reported new HIV/AIDS infections that were diagnosed during 2007.

Map 2.1. Location of laboratories and clinics that have reported data to HARS



2.2.5 Clinician reporting of HIV and AIDS diagnoses and deaths

In September 1982 clinicians began to confidentially report newly diagnosed cases of AIDS to the HPA[141]. Due to the progressive nature of infection without effective antiretroviral therapy, individuals are considered to be AIDS cases from the first diagnosis of AIDS-defining illness. The status of the patient does not revert with successful treatment of or spontaneous recovery from the illness.

The form used to report AIDS cases was revised with each revision of the AIDS case definition in 1985, 1987, 1989, 1993 and 1996 and additional minor changes were made in 1997 and 1998 to add further questions about antiretroviral therapy (Appendix Table B.2 and Appendix Forms B.2a – B.2g). These forms have also requested the date of death of HIV-infected patients who are known to have died. In January 2000, the AIDS report form was adapted to also allow clinicians to report newly diagnosed HIV infections (Appendix Table B.3 and Appendix Forms B.3a – B.3c). Clinicians are not requested to report HIV-infected individuals they see for the first time if they know them to have been diagnosed previously as these are expected to be reported by the laboratories if not by the previous clinician.

Since 1987, both AIDS forms and CHRs have requested that clinicians notify the HPA by telephone or letter of the date and cause of death of patients who die after being reported on an AIDS form or CHR.

2.2.6 ONS files of death certificates

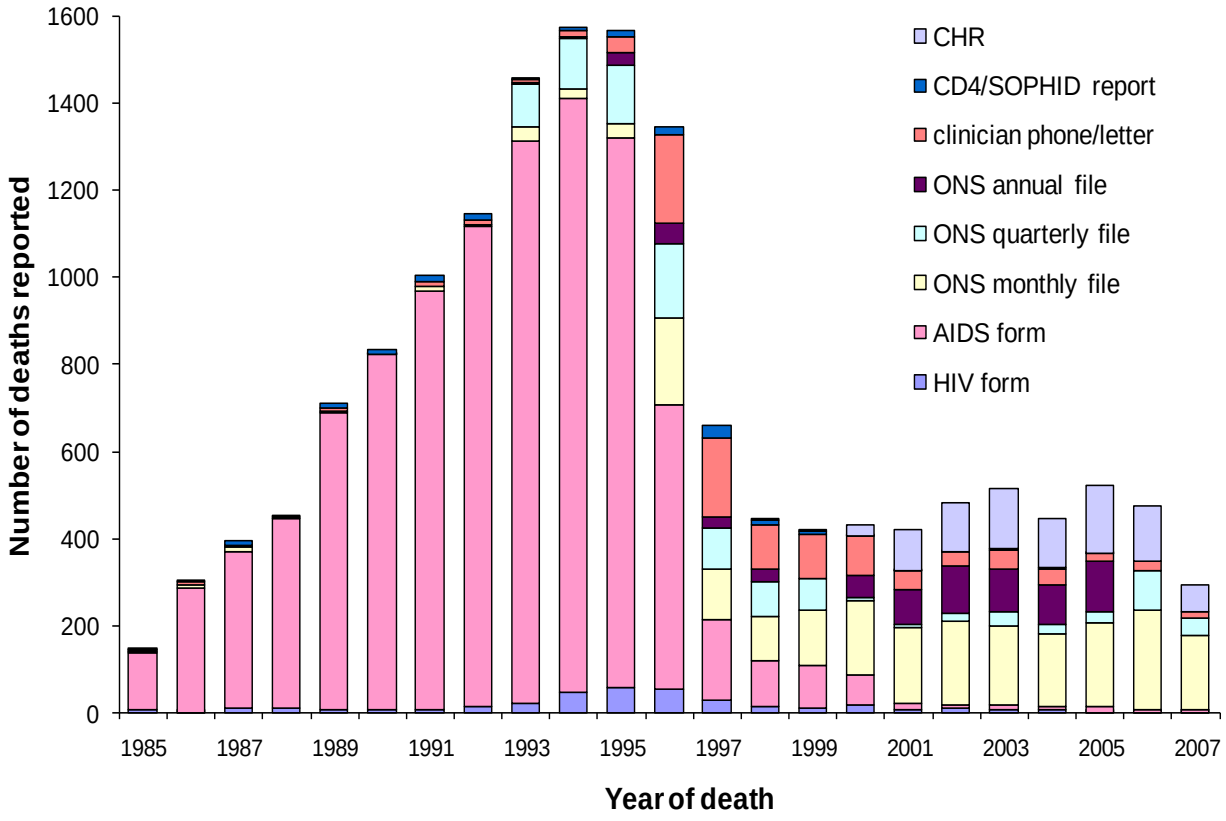
ONS has provided notifications of deaths occurring in HIV-infected individuals to complement the reporting of newly diagnosed HIV infections, since 1982. These add to the reports clinicians are asked to make when a patient dies, may identify some previously unreported cases (see below), and may provide causes of death not otherwise reported.

On a monthly, quarterly and annual basis, files of deaths are compared with HARS to see whether any further deaths in HIV-infected individuals can be identified. Monthly files, received since 1982, contain a report of every death notified in the previous month on which HIV, AIDS, Pneumocystis pneumonia, Kaposi's Sarcoma or conditions suggestive of these as a cause of death were recorded. Quarterly files contain notifications of all individuals aged less than 60 years who have died in the previous quarter to allow record linkage to deaths where HIV-related conditions have not been reported. Quarterly files have been received prospectively since 1997 but historical files were used at that time to update HPA records of deaths occurring between January 1993 and December 1996. Although this information is timely, not all deaths are included due to delayed notifications (for example, notifications awaiting a coroner's investigation). Therefore, annual files are also requested, which contain the final underlying cause of death information (which is not part of the public death certificate) for all deaths in those aged less than 60 years by calendar year of death. Restricting the data to those dying at age less than 60 years provides a manageable data set while including the majority of those who might have been

HIV-infected. Annual datasets are produced from the ONS 'Statistics' file on which their outputs are made; these are complete but not timely as this process is not finalised until autumn in the year following the date of death. Annual ONS files were first received in 2002 for individuals dying in 2000.

Figure 2.3 shows how quarterly and annual ONS files improved the ascertainment of deaths in 1993 and 2000 respectively.

Figure 2.3. Source of first received death report by year of death



A death record is added for any death report that can be identified as relating to an individual with both HIV and AIDS reports. Follow-up to ascertain whether the patient ever had AIDS is carried out for any death report that is linked to an

HIV report without a linked AIDS report. Where there is no link to existing reports but the ONS report indicates an HIV-related death or an illness usually associated with HIV, follow-up for both HIV and AIDS information is carried out. For follow-up, a physician at the site that reported the death is written to and asked to complete and return a CHR form (AIDS form before 2000). This confirms the HIV status and usually provides earlier dates of HIV and/or AIDS diagnoses. In recent years, about 100 further deaths have been assigned to patient records in HARS from annual ONS files.

2.2.7 Delayed reporting of HIV diagnoses, AIDS and deaths

A considerable number of reports of HIV diagnoses, AIDS diagnoses and deaths continue to be received months and years after the events occur. Just less than a third of the HIV diagnoses that were reported (by 2008) to have occurred in the first quarter of 2003 were received in the same quarter as the diagnosis occurred (Table 2.1). Almost 15% were received after more than one year and almost 5% were received after more than two years. The equivalent percentages for AIDS and deaths, respectively, were 21% and 18% reported after more than one year and 8% and 13% reported after more than two years. The greater percentage of death reports delayed by more than two years and the peak in the number of death reports delayed by 10 quarters reflects the timeliness of ONS files of deaths rather than clinician or laboratory reporting. Similar reporting delay affects surveillance systems in other countries[145].

Table 2.1. Numbers of HIV diagnoses, AIDS diagnoses and deaths (with or without AIDS) reported (by the end of 2007) to have occurred in the first quarter of 2003 by the number of quarters before reports were received

Number of quarters delayed	Number (%) of events that occurred in the first quarter of 2003		
	HIV diagnoses	AIDS diagnoses	Deaths
0	615 (31.4)	60 (26)	43 (31.2)
1	747 (38.1)	63 (27.3)	59 (42.8)
2	172 (8.8)	28 (12.1)	6 (4.3)
3	124 (6.3)	31 (13.4)	5 (3.6)
4	39 (2)	11 (4.8)	2 (1.4)
5	118 (6)	8 (3.5)	4 (2.9)
6	34 (1.7)	8 (3.5)	1 (0.7)
7	22 (1.1)	3 (1.3)	0 (0)
8	23 (1.2)	6 (2.6)	1 (0.7)
9	11 (0.6)	4 (1.7)	0 (0)
10	33 (1.7)	1 (0.4)	15 (10.9)
11	2 (0.1)	0 (0)	0 (0)
12	1 (0.1)	1 (0.4)	0 (0)
13	3 (0.2)	1 (0.4)	2 (1.4)
14	2 (0.1)	2 (0.9)	0 (0)
15	5 (0.3)	2 (0.9)	0 (0)
16	6 (0.3)	2 (0.9)	0 (0)
17	0 (0)	0 (0)	0 (0)
18	2 (0.1)	0 (0)	0 (0)
19	1 (0.1)	0 (0)	0 (0)
Total	**1,960 (100)**	**231 (100)**	**138 (100)**

2.2.8 Under-reporting of HIV and AIDS diagnoses and deaths

Under-reporting of HIV diagnoses, AIDS diagnoses and deaths to HARS is suspected because a number of records reported to CD4 Surveillance, SOPHID or ONS cannot be linked to HARS. Under-reporting of AIDS and deaths have also been frequently reported in the literature[146-156]. However, the degree of

under-reporting of these events has not been determined due to the difficulty in confirming whether the unlinked records represent different individuals or individuals that have already been reported to HARS but with different patient identifiers. Additionally, because of the passive nature of HARS reporting, it is difficult to ascertain whether these events would have been reported in the future to distinguish between delayed reporting and under-reporting. For deaths, doctors may be reluctant to record HIV on what is essentially a public document, reducing the number of deaths among individuals with a prior AIDS diagnosis that are identifiable from this source as HIV-infected[148;152;157-161]. In addition, some HIV cases may not be detected at death if the post-mortem did not indicate HIV infection. In particular, there were deaths that were not associated with an AIDS illness prior to the availability of effective combination therapy and non-AIDS related deaths have accounted for an increasing proportion of all deaths among HIV-infected individuals since then[152;162].

2.2.9 Follow-up of missing information

Follow-up is carried out for missing patient identifiers and missing CD4/viral load information. Clinics are telephoned for any missing data on the following:

- Soundex code, date of birth, sex (the few reports without these pieces of information have not been added to the database since January 2000);
- Any CD4 cell counts, CD4 cell count measurement dates, viral loads or viral load measurement dates missing from CHRs.

If the local patient ID number and/or risk group are missing then those data are also requested during follow-up for the above information.

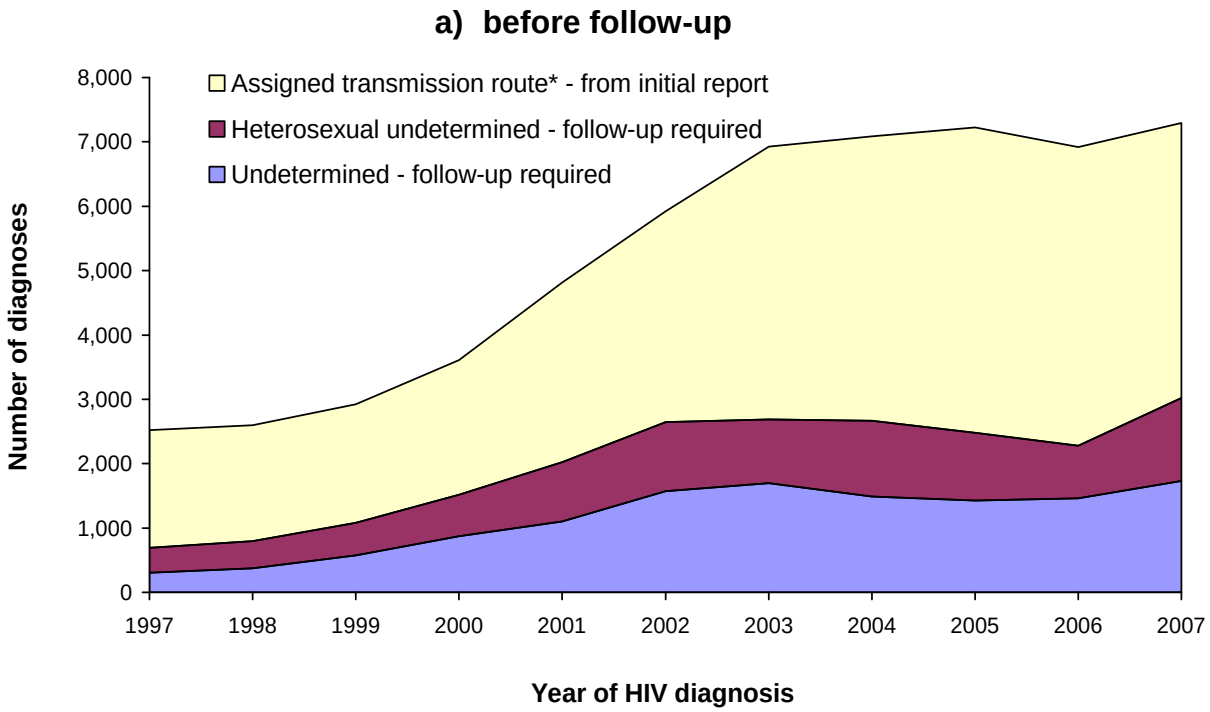
Follow-up is also carried out to reduce the proportion of records with missing route of infection data. Incomplete information makes determination of the number of new diagnoses acquired through uncommon routes difficult, may mask trends, and introduce selection bias to analyses.

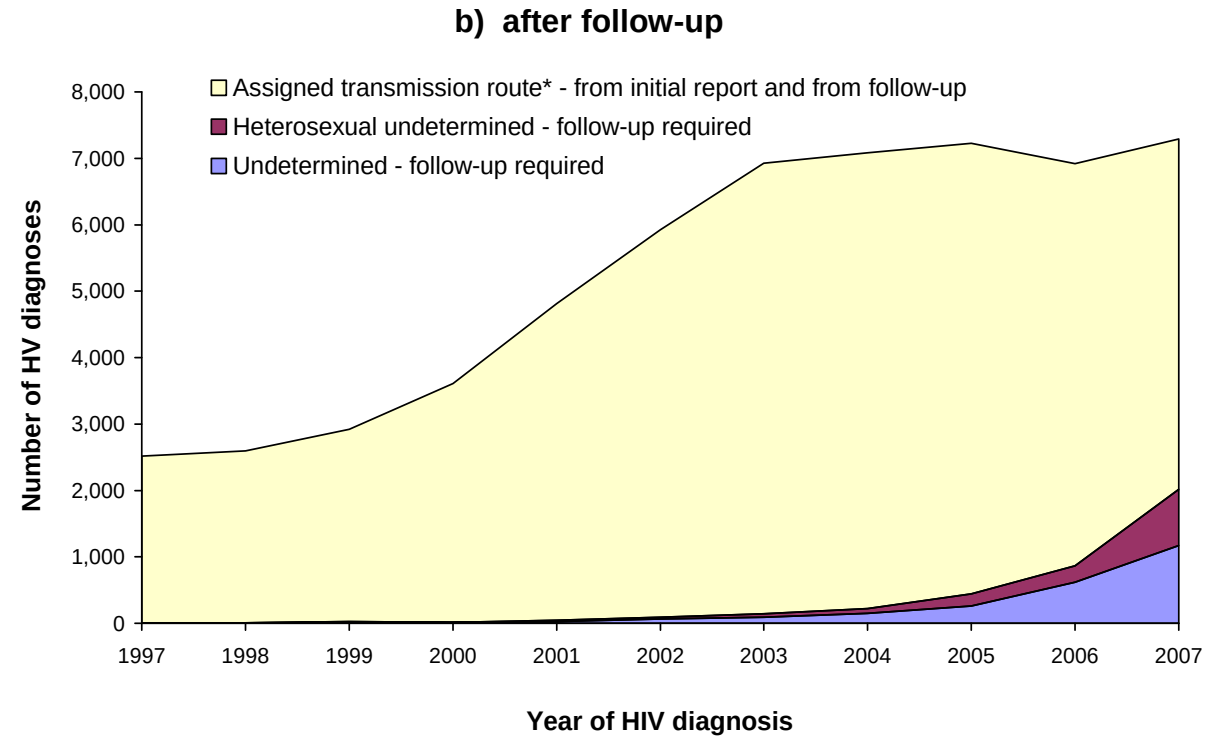
Follow-up is carried out for reports of all individuals who cannot initially be assigned to a route of infection category. This includes all those with no known route of infection (Figure 2.4 – category 'undetermined') and heterosexuals for whom information is missing on partner's risk group or likely country of infection (if the partner is presumed to have been heterosexually infected) (Figure 2.4 – category 'heterosexual undetermined')[163]. To monitor onward heterosexual HIV transmission in E,W&NI, follow-up is also carried out for clarification of risk group where it was reported to be heterosexual and where the sexual contact took place in the UK with a partner not known to be from a country where heterosexual transmission is common and not known to be at high risk of HIV infection[163].

The missing information is requested from microbiologists and/or clinicians or by data extraction from patient notes or databases at clinical centres where necessary. In-depth interviews with patients are sought, with the consent and facilitation of the clinician, when the likely route of transmission is otherwise unresolved. Any other missing information such as ethnicity is collected where possible if follow-up is undertaken for the reasons outlined above.

The graphs below show reports of new diagnoses in E,W&NI between 1997 and 2007, before and after follow-up. Follow-up was required for 12,607 reports with 'undetermined' route of infection and 9,223 reports with 'heterosexual undetermined' route of infection. Route of infection remained undetermined after follow-up for 24% and 17% of these records respectively. Follow-up was continuing for 17% (3,649) and abandoned for 2% (541) of cases. Cases were closed because patients had died (17%), patients were lost to follow-up (69%), clinicians advised against interviewing the patient (5%), patients refused interview (8%) or a route of transmission still could not be assigned after interview (1%).

Figure 2.4. Number of new HIV diagnoses in E,W&NI by risk group (a) before and (b) after follow-up





*includes sex between men, injecting drug users, all blood exposures, mother to infant and other.

2.2.10 Deduplication of reports from the same individual

Linking reports from the same individual is important to collate information about events reported on different forms and to limit duplication of records of the same individual in the database. These may be reports of different events that are received at different times or from different sources.

Deduplication occurs when a new report form is added to the database. Patient identifiers and demographics are compared using a multiplicative scoring system and likely duplicates are manually assessed by two users before reports are merged into a single consistent master patient record (Appendix A.8). Quarterly deduplication is also carried out on the whole database by identifying potential duplicate records for the same individual with matching dates of birth, sex and forename initial that may have been reported with different soundexes. The ‘master’ record in the patient table stores the demographic details from the report with the earliest date of HIV diagnosis and the earliest date of HIV diagnosis, AIDS and death. Records are not deduplicated if information is inconsistent unless follow-up confirms which data are correct.

2.3 The CD4 Surveillance Scheme

2.3.1 Background

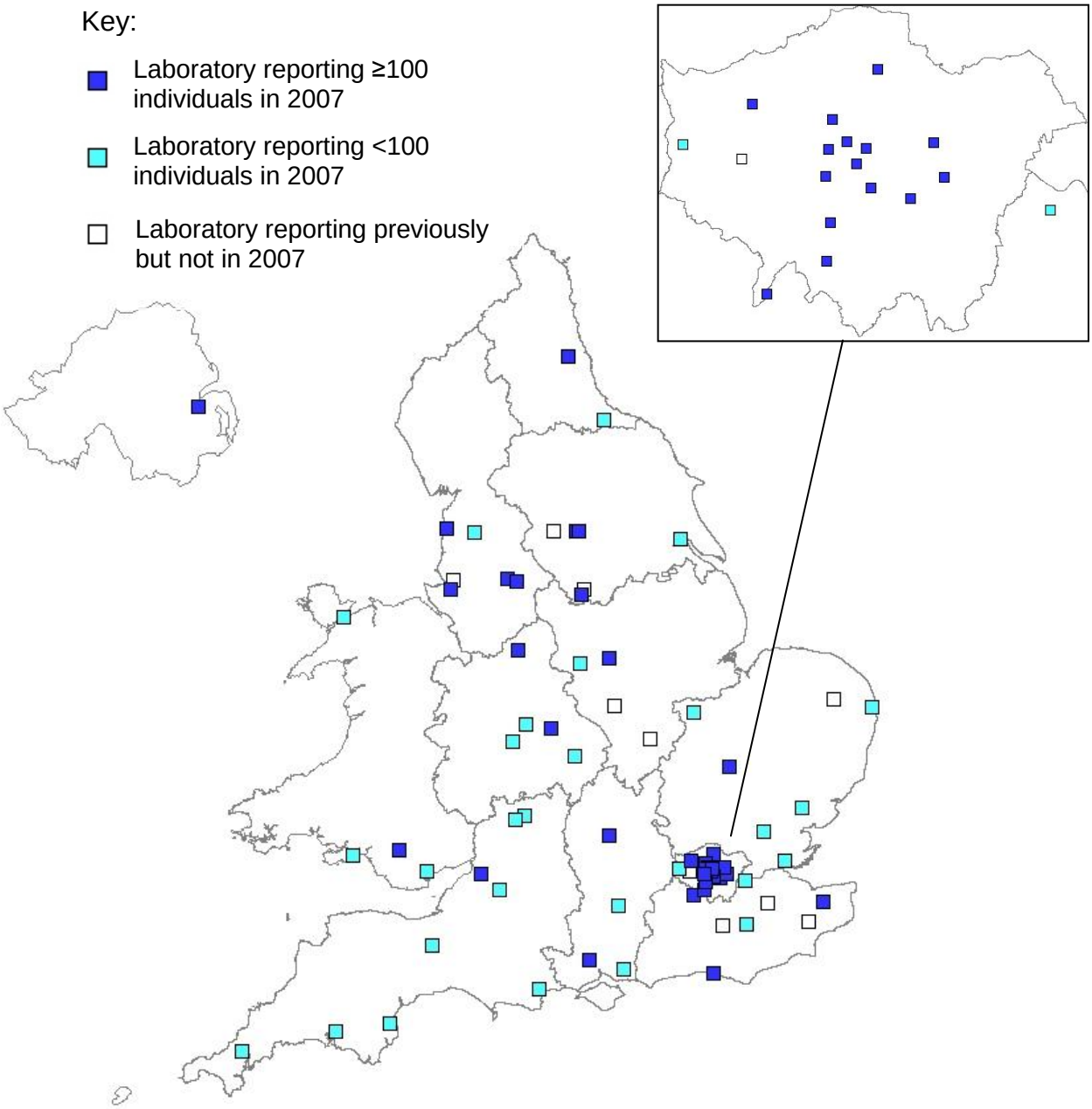
The national CD4 Surveillance Scheme was initiated in 1993 to augment the epidemiological information collected through the reporting of HIV/AIDS cases[164]. Laboratories performing CD4 cell counts and participating in the UK National External Quality Assessment Scheme are requested to provide all CD4 cell counts of HIV-infected adults (aged 15 years or more) to monitor trends in the prevalence of various levels of immunosuppression in E,W&NI.

2.3.2 Coverage of laboratories

Three laboratories initially reported CD4 cell counts in 1993 but by the end of 1996 50 laboratories had sent data to the scheme. On enrolment, laboratories were requested to send retrospective data to improve the completeness of the archive. A total of 72 laboratories had reported data by the end of 2007 although 10 of these no longer reported in 2007. Of these ten, seven had contracted this work to another laboratory included in the CD4 Surveillance Scheme; one was too busy, one had computing problems and one had staffing issues. Two laboratories reporting in 2007 could not provide data for 2005 or 2006 due to computing problems. An additional seven laboratories contacted had never sent data to the CD4 Surveillance Scheme; two were too busy, two did not respond, one only did clinical trials, one was not interested, and one was about to start reporting data in 2008. Incomplete data are usually available retrospectively in subsequent years.

The database therefore contains the majority of the CD4 cell counts performed on HIV positive individuals in E,W&NI between 1990 and 2007 and achieves substantial national coverage (Map 2.2).

Map 2.2. Geographic distribution of CD4 Surveillance Scheme participants

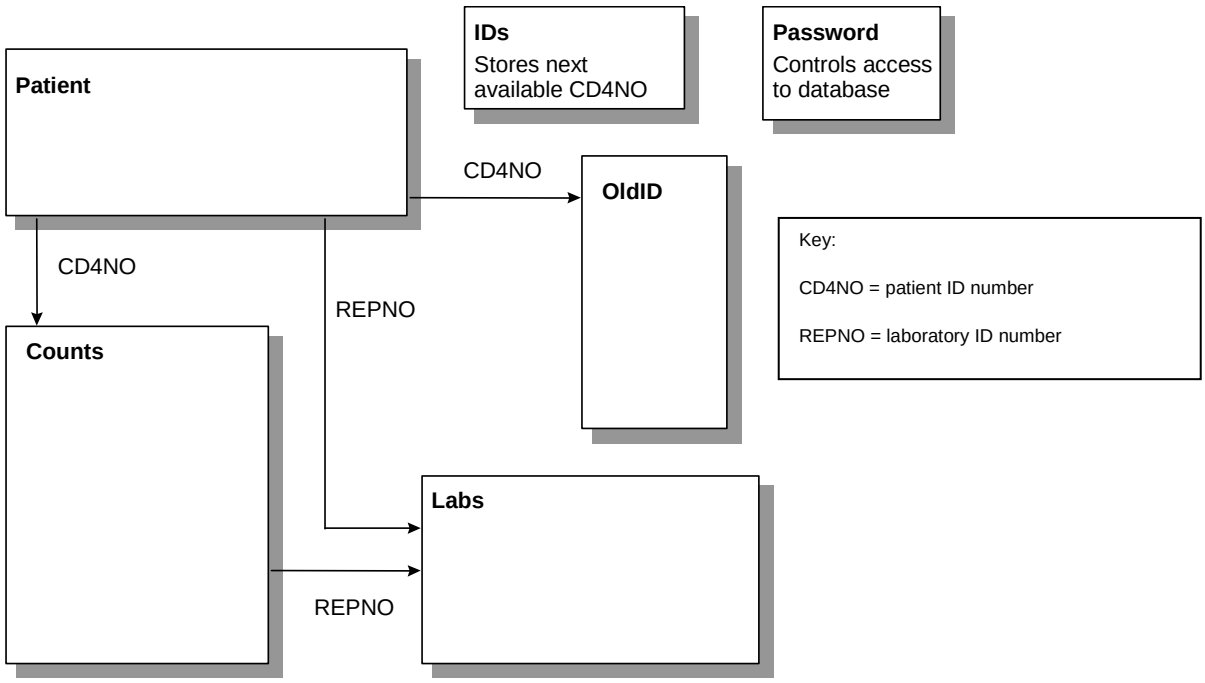


2.3.3 Data collected and data processing

Variables collected with each CD4 cell count include: soundex code, date of birth, initial, sex, primary local patient ID number, secondary local patient ID number (if provided), date CD4 cell count measured, absolute CD4 cell count, total lymphocyte count, percentage of lymphocytes that are CD4 cells, doctor requesting the CD4 count and the local clinic or hospital source of the sample. However, some laboratories cannot send soundex and date of birth information because they do not receive it with the sample to ensure confidentiality.

Information about individuals is used to populate a ‘patient’ table whereas the data about the CD4 cell counts is used to populate a ‘counts’ table (Figure 2.5). The patient table stores only the first reported set of demographic information for each patient regardless of whether that patient was subsequently reported with different demographics. Each record in the patient table is linked to as many records in the counts table as there are CD4 counts for that individual. CD4 cell counts are added to the database using the primary local patient ID number and reporting laboratory identifier. If the ID number matches an existing patient ID number on the database from the same laboratory, then the new counts are added to the counts table relating to that patient. The only change made to the patient table on addition of counts to an existing patient is to update missing information if that is present in the new report. If a patient number is reported that does not match an existing patient number on the database from the same laboratory then a new patient record is created and the new counts are added to the counts table relating to the new patient record.

Figure 2.5. Structure of the CD4 Surveillance database



CD4 Surveillance data are managed by a graduate scientist, led by a Masters-level epidemiologist, and overseen by a medical consultant epidemiologist.

2.3.4 Deduplication process

The process of importing data to CD4 Surveillance results in patients appearing more than once in the database if they are reported from different laboratories or with different patient ID numbers. This may occur if the patient has blood samples sent from different clinics to the same laboratory, if the patient attends a clinic that uses a different laboratory, if clinics change their administration systems, or if clerical errors result in a change in format of the laboratory identifier (e.g. M034746, 99M034746 and M34746). Therefore, every year after all of the previous years' counts have been received, the patient records are

deduplicated by matching patient identifiers such that all CD4 counts from duplicate records are attributed to a master record. The deduplication methodology was improved between 2004 and 2005 (Appendix A.9) based on the hierarchical algorithm developed for record linkage between the CD4 Surveillance and SOPHID databases (Section 3.6). In 2002, 1,664 records were deduplicated of which the majority had matching soundex and date of birth which probably identified individuals who had counts performed in different laboratories. In 2005, 7,561 records were deduplicated and the majority of these had missing or discordant soundex codes indicating greater sensitivity.

2.3.5 Record linkage to HARS

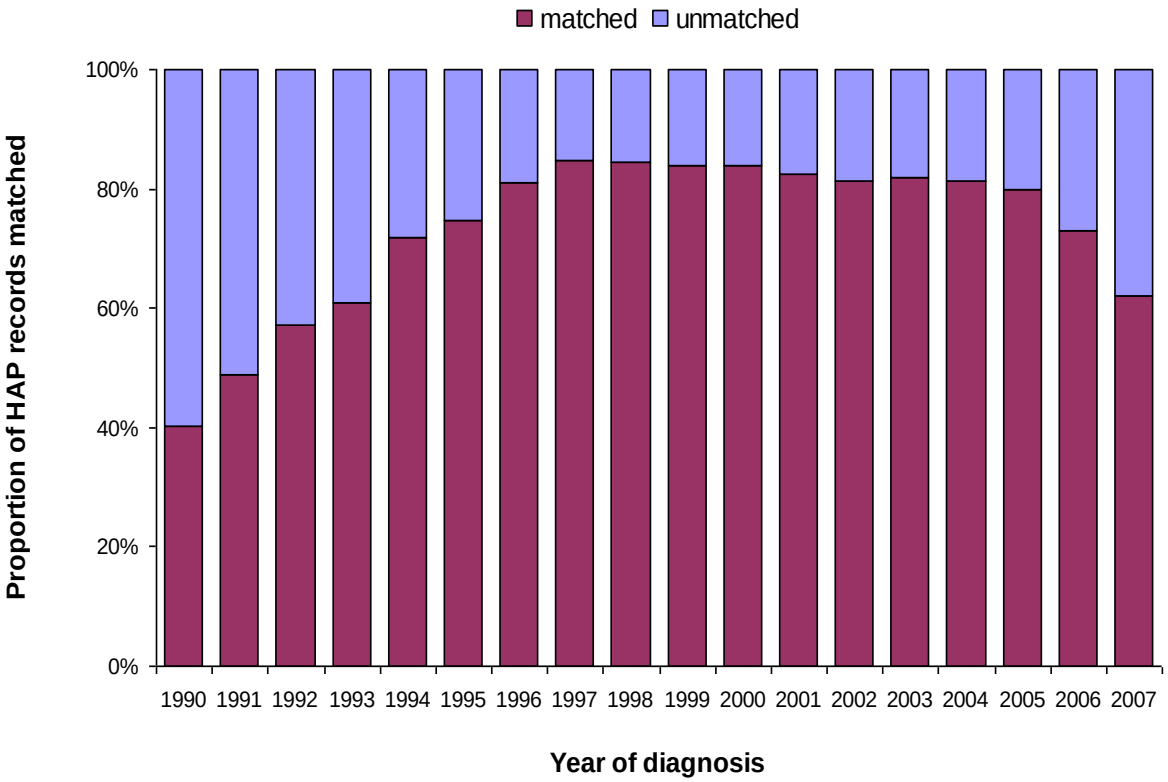
It is clear that some of the records for patients in CD4 Surveillance are duplicates of other patients or are from HIV-negative patients. By the end of 2007, there were a cumulative total of 1.3 million CD4 cell counts from over 104,761 adults on the database. However, only 87,193 HIV-infected adults (aged 15 years or more) had been reported as newly diagnosed in E,W&NI to HARS by June 2008. Although laboratories are asked to report only CD4 cell counts from HIV-infected adults, some do not exclusively perform CD4 cell counts on HIV-infected individuals and do not have the resources or information to exclude CD4 cell counts of uninfected individuals from surveillance data.

To ensure that CD4 counts analysed are from HIV-infected individuals, record linkage is carried out annually between the CD4 and HARS databases (Appendix A.10). This process also relates CD4 counts to data such as date of HIV diagnosis, ethnicity, date of AIDS diagnosis, and risk group.

Record linkage related about four in five of records in HARS to records in the CD4 Surveillance database from the mid-1990s when national coverage of CD4 Surveillance was achieved (Figure 2.6). Incomplete linkage may have been due to individuals reported to HARS not having CD4 cell counts, having CD4 counts that were not reported to CD4 Surveillance or having CD4 counts that were reported to CD4 Surveillance but with patient identifiers that did not match.

Individuals with positive HIV test results may not have had CD4 cell counts if they did not return for their results or were lost to follow-up once they received their result. However, this was likely to account for only a minority of the non-linked reports because most people are likely to return for care as clinical symptoms become increasingly significant. Only a relatively small number of people were likely to have died or emigrated before returning for care[165;166].

Figure 2.6. Proportion of HARS records linked to CD4 records (Sept. 2008)



Some of the incomplete linkage was likely to be due to CD4 cell counts not having been reported to CD4 Surveillance. However, between 1998 and 2007, there were only 620 reports of new HIV diagnoses from HARS clinics that related to laboratories that had never reported to the CD4 Surveillance Scheme. These accounted for only 1.1% (378/55,338) of reports of HIV diagnoses.

Different or incomplete reporting of patient identifiers was therefore likely to account for the majority of non-linkage. Although there were very few reports of new HIV diagnoses with missing patient identifiers during this period, this was relatively common in reports to CD4 Surveillance (about one in four soundex codes and one in twelve dates of birth). It was clear that a minority of reports that linked between the databases had different patient identifiers, which was assumed to be largely due to coding errors either locally or at HPA. It was assumed that only a few reports would have even more dissimilar patient identifiers in the HARS and CD4 Surveillance databases as the likelihood of these, more extensive coding errors, would be even lower. Therefore, it was assumed that most individuals reported to both surveillance systems but not linked were due to missing soundex code or date of birth in CD4 Surveillance.

Each year, after record linkage, the proportion linked was consistently lower for the most recent years. This may be due to delayed reporting of HIV diagnoses (Section 2.2.7) or the temporary loss to follow-up of some recently-diagnosed patients as they take time to come to terms with their diagnosis[167-171]. Patients diagnosed towards the end of the most recent calendar year may not have had their first CD4 cell counts until the next calendar year.

2.4 The Survey of Prevalent HIV Infections Diagnosed

2.4.1 Background

The Survey of Prevalent HIV Infections Diagnosed (SOPHID) has carried out annual cross-sectional surveys of every individual who attended for HIV-related care at NHS hospitals in E,W&NI during the previous calendar year since 1995[172]. The objective of the SOPHID surveys is to provide accurate information on the current number and distribution of individuals aged 15 years or more who have received HIV-related care from statutory services for use in funding allocation, health-care planning and health promotion. The key aspect of these data is the current area of residence (Appendix A.11) of the individuals reported. The Department of Health for England originally funded the national survey for allocation of ring-fenced HIV funding although this information has been increasingly used locally for residence-based cross-charging by HIV commissioners. This information cannot be obtained from reports of new HIV/AIDS diagnoses because patients' place of residence and treatment often change after diagnosis[173]. A key aspect of the SOPHID surveys is that data collection does not finish until all reporters have submitted and validated data – therefore, there is no delayed reporting and records from previous years are not updated.

2.4.2 Methodology

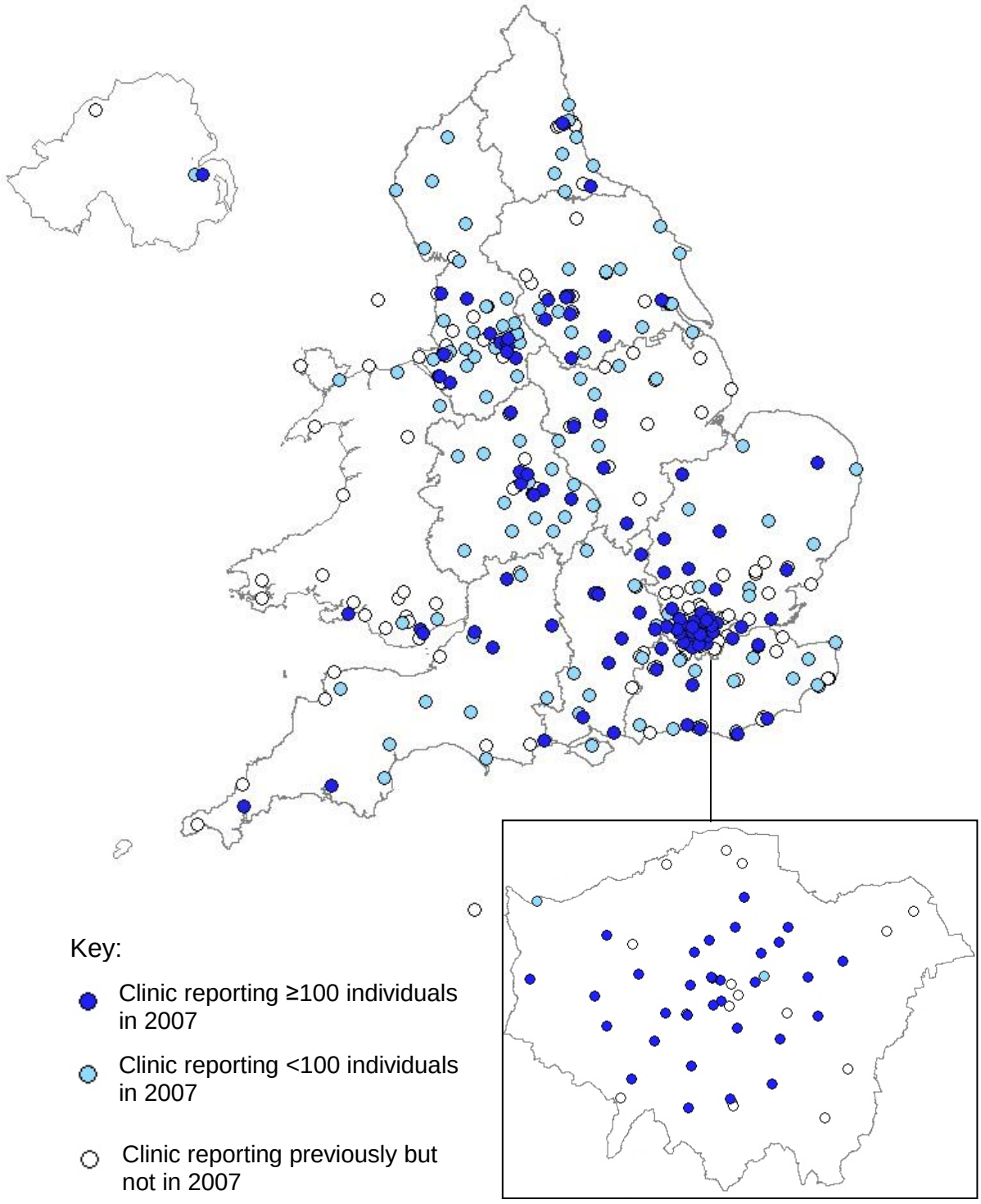
Local co-ordinators collate data (Appendix A.12) for their area/hospital/clinic (Map 2.3) and send it to the HPA. Originally, one co-ordinator, with a good knowledge of the local area, was identified in each Health Authority (Appendix

A.11) but progressively data providers have been identified for each reporting hospital/clinic to improve the communication process. Overall, 332 clinics in E,W&NI had reported to SOPHID to the end of 2007 with 246 of those reporting in 2007 (Map 2.3). In London and East Sussex (referred to as the London survey) the SOPHID surveys have been carried out biannually since 1998 and in other areas (referred to as the national survey) once a year.

SOPHID data are managed by two graduate scientists, two Masters-level epidemiologists, and overseen by a medical consultant epidemiologist.

The collection of information used to link reports from the same patient within and between surveys have always included soundex code of the surname, date of birth and sex. First name initial has been requested since 2000 in the national surveys and since 2002 in the London surveys to help clarify links between records. Local patient ID number has been collected for individuals seen for care in London since the start of the London SOPHID surveys in 1998 but has only been requested outside London since 2003. This was due in part to recognition that soundex codes and dates of birth were sometimes reported differently for the same individual (due to either clerical errors or misunderstanding/miscalculation of the soundex algorithm) and were indeed expected for the majority of women who married and changed their surnames (Chapter 3).

Map 2.3. Geographic distribution of hospitals/clinics that reported data to SOPHID



Collection of information about the stage of HIV infection developed over time. CD4 cell count was added to the data collection request in 2000. 'Vital status' (either alive or dead at the end of the year) was collected until 1999 but since 2000, most advanced clinical stage has been collected in five categories of: asymptomatic, symptoms pre-AIDS, AIDS, death in a patient with AIDS, and death in a patient who has never had AIDS.

The national and London surveys were standardised and new variables were added in 2005 after I became responsible for managing the SOPHID surveys in May 2005. Collection of Health Authority of residence in the London survey was discontinued at this time. Datepos was added to the data request in the national survey (it was requested since the start of the London surveys in 1998). However, its definition was changed from the first ever date of a positive HIV test or HIV diagnosis to the first date positive or date of first attendance at that site. This was to reduce duplication between the reporting requests of new diagnoses and individuals accessing care and to provide information on transfers between sites. The new variables requested were latest viral load, viral load date, previous care, date of starting ART and date of AIDS diagnosis (since the second half of 2005 in the London surveys).

2.4.3 Data processing, validation and follow-up

SOPHID data were originally requested to be sent to the HPA on paper forms for fewer than 20 reports, or on diskette if numbers were greater. Paper reports were manually entered into a single flat file database table by a single user. Progressively, data have been sent in Excel spreadsheets and since 2006, the

spreadsheet has included integral validation for each variable. This provides pop-up messages to the user indicating the format of the information required, and pop-up error messages if the data entered are inconsistent with the expected formatting (e.g. numbers in text fields, dates not in dd/mm/yyyy format). The pop-up error messages appear on manual data entry. If data are copied and pasted into the spreadsheet, a macro circles in red all data that are inconsistent with the expected formatting for attention prior to submission of the dataset. The data are also re-validated on entry to the national database and up to two follow-ups are made with the local co-ordinators to minimise missing or inconsistent data if necessary.

2.4.4 Deduplication

Although data providers are asked to report each of their patients only once to each survey, deduplication is necessary for two reasons. Firstly, they may not have the resources to do that locally, and secondly patients may visit more than one site within a survey period. Data for the London surveys are deduplicated at the end of each six-month data-collection period and are used to produce epidemiological information for participants. London datasets from the same calendar year are amalgamated and deduplicated within reporting sites (those with soundex, date of birth, sex and SOPHID site matching) to leave the last attendance in the year at each site of care. This dataset is then appended to the national survey for the same calendar year and the combined dataset is subsequently deduplicated, such that only the last attendance during the calendar year for each individual remains in the annual dataset. Records are

considered to be potential duplicates from the same patient if the soundex code, date of birth and sex match. Additional data fields such as infection route, ethnicity, and postcode are then considered to prevent aggregation of data from truly different individuals. If records are found to be duplicates then the record with the latest date seen for care is retained while the other is excluded. Information provided in the excluded record that is missing from the retained record is used to update the record that is kept.

2.4.5 Amalgamated SOPHID dataset

A longitudinal dataset was created (initially by Brian Rice) by amalgamating the annual cross-sectional surveys from 1997 to 2002 and assigning an individual patient ID number to all records. Records from different survey years were assumed to originate from the same patient and linked if the soundex code, date of birth and sex matched. Where a combination of soundex code, date of birth and sex occurred more than once in a survey year indicating truly different individuals these were deduplicated manually to prevent incorrect linkage.

2.4.6 SOPHID patient and report tables

I devised and supervised the creation of a SOPHID ‘report table’ by appending together in one dataset all original files of reported data. This contained all the variables that were in the SOPHID reports as well a patient-specific number relating records from the same individual. This patient number was assigned according to criteria in a record linkage algorithm that required the matching of at least two of the three main patient identifiers (soundex, date of birth or local patient ID number) and then considered matches using the full postcode or

postcode sector+ (all but the last character of the postcode) (Table 2.2). All records that matched on any of the criteria shown were linked and considered to relate to an individual. Because of the high specificity expected from matching with local patient ID number and to allow for some coding errors, year of birth was used instead of date of birth in two of the matching criteria.

Table 2.2. Matching criteria in the record linkage algorithm for de-duplication of SOPHID records to create the SOPHID report table (✓ = exact match)

Match score	Postcode	Clinic	Local patient ID	Soundex	Date of birth	Sex
1				✓	✓	✓
2		✓	✓		Year of birth	
3		✓	✓	✓		
4	Full postcode				✓	
5	Postcode sector+			First three characters	Year of birth	✓

The SOPHID report table in 2008 contained 574,913 records from 1995 to 2007, of which 562,130 were from 79,596 adults (aged 15 years or more) according to the record linkage algorithm. If the same data were deduplicated according solely to exact soundex, date of birth and sex matches then it would have appeared that there were 89,076 adults overall and more in each year (Appendix A.13).

In the SOPHID patient table, the last reported soundex code, date of birth and sex are assigned to the patient record to reflect the most up-to-date situation and because information which changes for an individual between surveys are validated with the reporting clinic.

2.5 HIV epidemiology from case reporting

2.5.1 Overview

HARS data were originally collected to directly inform public health policy about transmission, as reflected by new diagnoses of HIV infections, and about the resource requirements for health services, as reflected by AIDS diagnoses. The characteristics of these can differ due to the long incubation period of infection.

A total of 95,627 HIV infections were newly diagnosed in the UK by the end of December 2007 (reported by the end of December 2008). This figure includes some records of the same individuals who could not be deduplicated because of differences in the information supplied, some records of individuals who left the country at some date after diagnosis and many records of individuals who have died. The total number of individuals reported to SOPHID as seen for care in the UK in 2007 was 56,377. Estimates including undiagnosed HIV infections suggest that there were a total of 77,400 individuals living with HIV in the UK in 2007. This equated to an overall HIV prevalence of 127 per 100,000 population.

The HIV epidemic in the UK is not equally distributed throughout the population but focused in risk groups (Section 1.3.5). Different epidemic trends have been observed in each group. The overall breakdown, to the end of December 2007, was as follows:

- 44.5% (42,525) sex between men
- 43.5% (41,619) heterosexual sex
- 5.1% (4,876) IDU

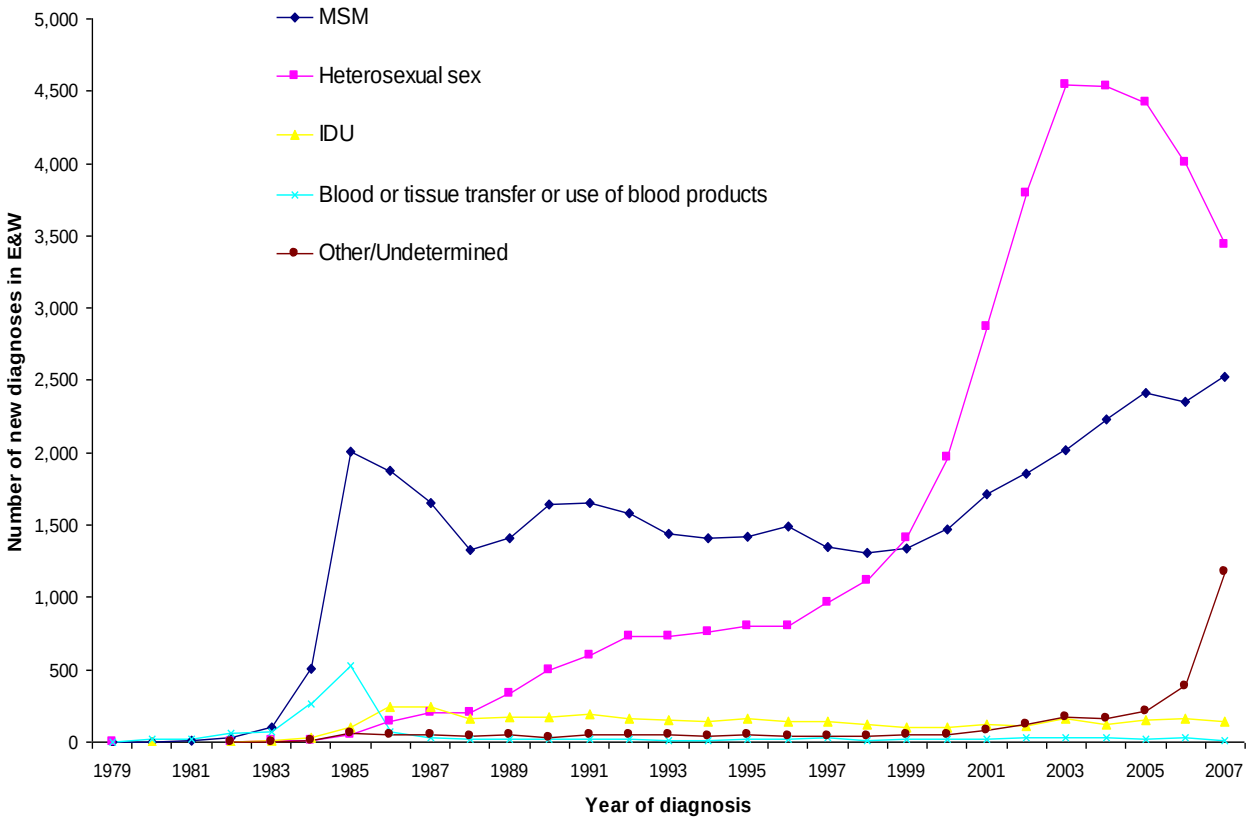
- 2.0% (1,893) blood or tissue transfer or use of blood products
- 1.8% (1,696) transmission from mother-to-infant
- 3.2% (3,018) other or undetermined

The geographical distribution of HIV-infected individuals within the UK is heterogeneous. The focus of the epidemic, in each exposure category, is London, which accounted for 49.2% (27,741/56,377) of all reported diagnosed HIV-infected UK residents when last seen for care in 2007.

2.5.2 Trends in new HIV diagnoses in E,W&NI reported to HARS

The earliest evidence of HIV infections diagnosed in E,W&NI comes from retrospective samples stored in 1979. The introduction of laboratory tests for HIV in late 1984 resulted in a rapid rise in the number of new diagnoses as asymptomatic, prevalent infections were detected. These infections were mostly due to sex between men and transmission due to the transfer of blood, tissues and blood products (Figure 2.7). Following the initial diagnosis of infections from the prevalent pool, the number of newly diagnosed infections acquired through sex between men was relatively stable until 2000 (around 1,500 per year) but then increased to a record 2,526 in 2007. There was a sharp decline in newly diagnosed infections acquired from blood, tissues and blood products after 1985 largely associated with the introduction of methods to screen and heat treat blood factor. More recent diagnoses due to these sources are most likely infections acquired in countries where these sources are less safe. Numbers of newly diagnosed infections acquired through IDU declined from 166 in 1988 to 98 in 2000 but then increased to previous levels by 2007.

Figure 2.7. The number of newly diagnosed HIV infections among adults (aged 15 years or above) by year of diagnosis and risk group: reported in E,W&NI by June 2008.



It is important to note that diagnoses often do not equate to incident infections

Numbers acquired through heterosexual infection increased to 733 in 1992, stabilised until 1996, and then rapidly increased to peak at 4,544 in 2003 before falling off to 3,435 in 2007. It is important to note that the number of new diagnoses, particularly for recent years, will rise as further reports are received. There were few diagnoses each year attributed to other routes of infection or that remained undetermined. These numbers for recent years will fall as follow-up continues.

The breakdown of newly diagnosed infections acquired heterosexually shows that the vast majority of individuals acquired their infection from partners who themselves were presumed to have been infected heterosexually (Table 2.3). The relatively small number of diagnoses among heterosexuals infected by a high-risk partner did not fall between 1990 and 2006 but accounted for a decreasing proportion of newly diagnosed infections acquired heterosexually.

Table 2.3. The number of HIV infections acquired heterosexually by year of diagnosis and heterosexual sub-category: reported in E,W&NI by June 2008.

Year of diagnosis	Heterosexual sub-category				Total
	High-risk partner	Heterosexual partner	Partner exposure undetermined		
			investigation continuing	investigation closed	
prior to 1990	206 (22%)	742 (78%)	0 (0%)	9 (1%)	**957**
1990	51 (10%)	445 (89%)	1 (0%)	3 (1%)	**500**
1991	54 (9%)	545 (91%)	0 (0%)	3 (0%)	**602**
1992	69 (9%)	659 (90%)	2 (0%)	3 (0%)	**733**
1993	66 (9%)	661 (91%)	0 (0%)	2 (0%)	**729**
1994	55 (7%)	698 (92%)	1 (0%)	4 (1%)	**758**
1995	57 (7%)	737 (92%)	3 (0%)	5 (1%)	**802**
1996	54 (7%)	739 (92%)	3 (0%)	3 (0%)	**799**

1997	67 (7%)	885 (92%)	3 (0%)	8 (1%)	**963**
1998	68 (6%)	1,043 (93%)	3 (0%)	3 (0%)	**1,117**
1999	42 (3%)	1,346 (96%)	15 (1%)	6 (0%)	**1,409**
2000	40 (2%)	1,916 (97%)	7 (0%)	9 (0%)	**1,972**
2001	70 (2%)	2,772 (97%)	24 (1%)	6 (0%)	**2,872**
2002	63 (2%)	3,681 (97%)	44 (1%)	8 (0%)	**3,796**
2003	60 (1%)	4,403 (97%)	75 (2%)	6 (0%)	**4,544**
2004	59 (1%)	4,339 (96%)	128 (3%)	11 (0%)	**4,537**
2005	65 (1%)	4,128 (93%)	224 (5%)	9 (0%)	**4,426**
2006	54 (1%)	3,493 (87%)	447 (11%)	8 (0%)	**4,002**
2007	32 (1%)	2,497 (73%)	902 (26%)	4 (0%)	**3,435**
Total	**1,232 (3%)**	**35,729 (92%)**	**1,882 (5%)**	**110 (0%)**	**38,953**

Further analysis of newly diagnosed infections acquired heterosexually from partners who themselves were infected heterosexually shows that the majority had had sexual partners in Africa (Table 2.4). It has been shown that these sexual exposures were mainly occurring in Uganda and Kenya in the early 1990s, but that although the number of infections attributed to Uganda and Kenya increased to 2002, Zimbabwe was the source of the majority of exposures in more recent years[174]. Data on year of arrival, collected since January 2000, have also shown that there was a sharp peak in 2002 in the number of individuals infected with HIV in Africa who arrived in the UK.

Evidence that the majority of heterosexually-infected individuals diagnosed in the UK acquired their infection abroad means that sex between men is the most common route of acquisition of HIV in the UK[115].

Table 2.4. The number of HIV infections acquired heterosexually from heterosexuals by year of diagnosis and sub-category: reported in E,W&NI by June 2008.

Year of diagnosis	Region of exposure									
	Exposure abroad							Exposure in the UK		
	Africa	Latin America / Caribbean	Asia	North America	Europe	Australasia	country(ies) not known	outside Europe	within Europe	country(ies) not known
prior to 1990	554 (75%)	16 (2%)	18 (2%)	25 (3%)	38 (5%)	3 (0%)	18 (2%)	36 (5%)	30 (4%)	4 (1%)
1990	361 (81%)	11 (2%)	5 (1%)	7 (2%)	23 (5%)	1 (0%)	3 (1%)	14 (3%)	19 (4%)	1 (0%)
1991	437 (80%)	12 (2%)	16 (3%)	10 (2%)	22 (4%)	1 (0%)	0 (0%)	16 (3%)	31 (6%)	0 (0%)
1992	511 (78%)	23 (3%)	23 (3%)	14 (2%)	34 (5%)	1 (0%)	1 (0%)	27 (4%)	25 (4%)	0 (0%)
1993	497 (75%)	24 (4%)	28 (4%)	16 (2%)	32 (5%)	2 (0%)	0 (0%)	19 (3%)	41 (6%)	2 (0%)
1994	526 (75%)	26 (4%)	19 (3%)	9 (1%)	36 (5%)	0 (0%)	0 (0%)	40 (6%)	42 (5%)	0 (0%)
1995	543 (74%)	13 (2%)	35 (5%)	8 (1%)	43 (6%)	2 (0%)	2 (0%)	50 (7%)	36 (5%)	5 (1%)
1996	544 (74%)	24 (3%)	44 (6%)	7 (1%)	40 (5%)	1 (0%)	6 (1%)	42 (6%)	31 (4%)	0 (0%)
1997	635 (72%)	31 (4%)	46 (5%)	9 (1%)	45 (5%)	2 (0%)	3 (0%)	75 (8%)	38 (4%)	1 (0%)
1998	741 (71%)	32 (3%)	72 (7%)	14 (1%)	38 (4%)	3 (0%)	14 (1%)	88 (8%)	40 (4%)	1 (0%)
1999	997 (74%)	69 (5%)	71 (5%)	7 (1%)	49 (4%)	5 (0%)	0 (0%)	99 (7%)	48 (4%)	1 (0%)
2000	1,489 (78%)	72 (4%)	110 (6%)	8 (0%)	42 (2%)	2 (0%)	2 (0%)	139 (7%)	51 (3%)	1 (0%)
2001	2,274 (82%)	97 (3%)	94 (3%)	8 (0%)	48 (2%)	4 (0%)	0 (0%)	184 (7%)	51 (2%)	12 (0%)
2002	3,028 (82%)	148 (4%)	123 (3%)	7 (0%)	67 (2%)	4 (0%)	0 (0%)	258 (7%)	35 (1%)	11 (0%)
2003	3,592 (82%)	168 (4%)	143 (3%)	7 (0%)	93 (2%)	5 (0%)	1 (0%)	317 (7%)	69 (2%)	8 (0%)
2004	3,434 (79%)	148 (3%)	174 (4%)	10 (0%)	95 (2%)	0 (0%)	1 (0%)	385 (9%)	63 (1%)	29 (1%)
2005	3,183 (77%)	106 (3%)	177 (4%)	15 (0%)	93 (2%)	3 (0%)	3 (0%)	415 (10%)	81 (2%)	52 (1%)
2006	2,551 (73%)	92 (2%)	176 (5%)	9 (0%)	113 (3%)	1 (0%)	1 (0%)	374 (11%)	88 (3%)	88 (3%)
2007	1,716 (69%)	48 (2%)	131 (5%)	10 (0%)	86 (3%)	2 (0%)	0 (0%)	318 (13%)	80 (3%)	106 (4%)
Total	**27,613 (77%)**	**1,160 (3%)**	**1,505 (4%)**	**200 (1%)**	**1,037 (3%)**	**42 (0%)**	**55 (0%)**	**2,896 (8%)**	**899 (3%)**	**322 (1%)**

2.5.3 Trends in HIV diagnoses and deaths in E,W&NI reported to HARS by risk group

Since 1990, the overall numbers of new HIV diagnoses and deaths have been dominated by infections acquired sexually. From 1990 to 1995 the number of deaths increased while the number of newly diagnosed infections remained relatively stable. The steep rise in the number of newly diagnosed, sexually-acquired infections and a dramatic fall in the number of deaths among MSM (Figure 2.8) resulted in rapidly increasing numbers of patients accessing care after 1995.

2.5.4 Trends in AIDS diagnoses and deaths in E,W&NI reported to HARS

HARS only captures the first diagnosis of AIDS for each individual. Trends in the number of first AIDS diagnoses reported to HARS are similar to trends in deaths. Following gradual increases in the number of AIDS diagnoses between the early 1980s and mid-1990s, there were marked decreases as HAART became widely available (Figure 2.9). However, there was an increase in the number of first AIDS diagnoses reported between 2001 and 2003.

Figure 2.8. Number of new diagnoses by year of diagnosis and deaths by year of death: reported in E,W&NI by June 2008.

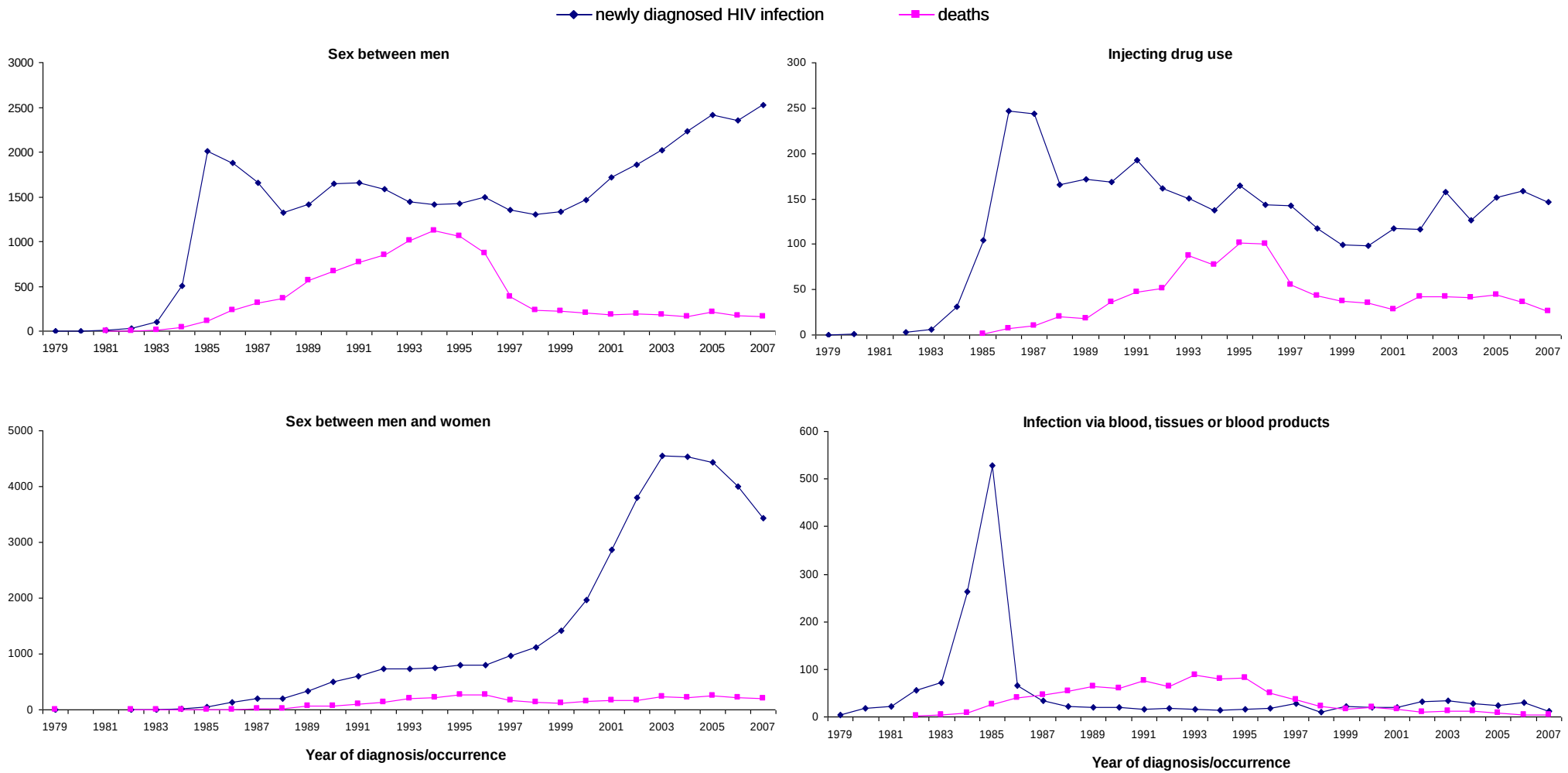
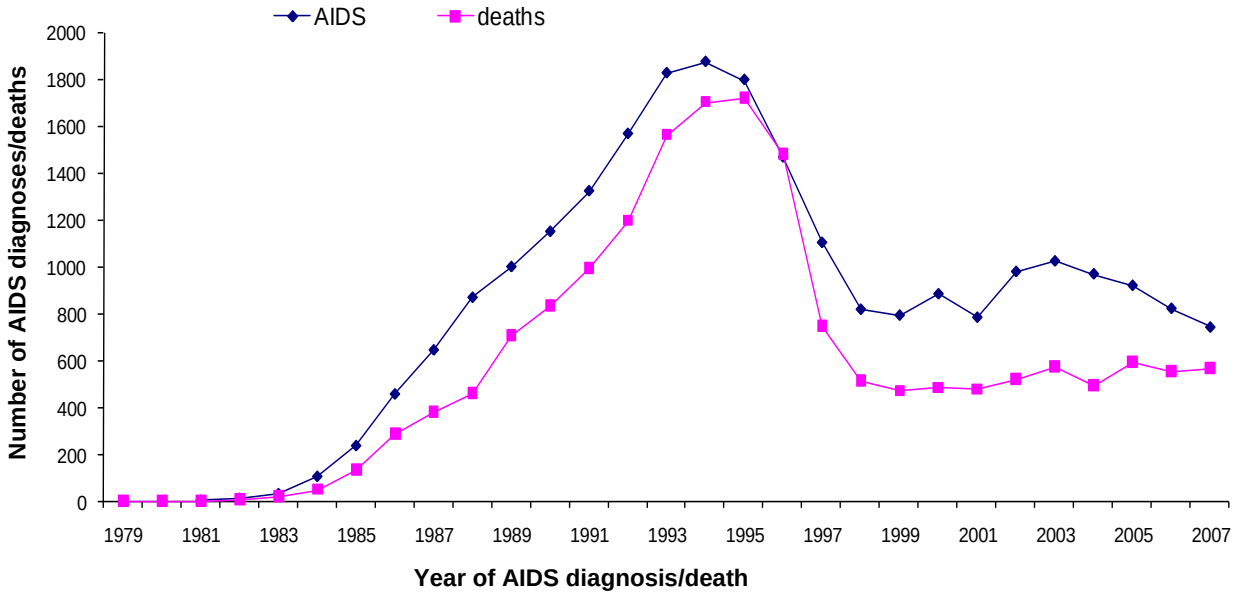


Figure 2.9. Trends in the number of first AIDS diagnoses by year of diagnosis and deaths by year of death.



2.5.5 Number of diagnosed HIV-infected individuals accessing care in E,W&NI reported to SOPHID

There were 53,673 diagnosed HIV-infected patients seen for care in E,W&NI in 2007. Sex between men accounted for 42.7% (22,935) of these infections, sex between men and women for 50.5% (27,131), injecting drug use 2.1% (1,111) and other/unknown exposures, 4.7% (2,496).

Almost all (99.1% [53,180]) individuals accessing care in 2007 had a reported ethnic group. Of those who acquired HIV infection through sex between men, 87.5% were white, 7.2% were of Asian, Oriental, Indian/Pakistani/Bangladeshi or mixed ethnicity and 4.5% were of black ethnicity. In contrast, the ethnic breakdown of individuals heterosexually infected was 68.4% of black African ethnicity, 18.7% white, 6.0% Asian, Oriental, Indian/Pakistani/Bangladeshi or mixed ethnicity and 6.4% of black Caribbean or other black ethnicity.

In 2007, 71.1% (37,699/52,997) of individuals were on therapy (excluding 676 individuals without a reported level of antiretroviral therapy). The number of individuals seen for care in E,W&NI increased by 9.0% between 2006 and 2007 (from 49,219) and nearly tripled between 1999 and 2007 (from 19,127).

2.6 HIV case report surveillance in other countries with similar epidemics

HIV case report surveillance in other countries with low-level or concentrated epidemics is based on similar principles to that in E,W&NI related to the nature of the epidemic. Case-based surveillance systems originated for the reporting of clinical AIDS diagnoses. Most of these systems were developed by addition of reports of HIV diagnoses and some furthermore by monitoring laboratory data such as CD4 cell counts and/or viral loads. The SOPHID methodology of repeated cross-sectional surveys of prevalent, diagnosed HIV-infected individuals accessing care is unique but equivalent information can be derived from longitudinal or cohort data.

The mechanisms of reporting (including whether voluntary or mandatory), the degree of integration of systems, the range of data variables collected and the national coverage of surveillance vary substantially[175-178]. Yet, HIV case report surveillance can be broadly grouped into five categories based upon these factors: i) systems with national coverage of patient-specific HIV and AIDS case reporting, which may also have introduced the monitoring of laboratory markers such as in Scotland, E,W&NI and Australia; ii) systems with national coverage that were limited to AIDS case reporting until the recent introduction of HIV case

reports such as in France, the United States and Ireland; iii) systems with national coverage of patient-specific AIDS case reporting but anonymous reporting of HIV cases such as in Denmark and Germany; iv) systems with national coverage of patient-specific AIDS case reporting but no national HIV case reporting such as in Italy and Spain (France before 1998 and the United States before 2008); v) national cohorts such as in the Netherlands and Austria.

HIV surveillance systems in Scotland are similar to those in E,W&NI as described earlier. Reporting of CD4 cell counts from laboratories was added to HIV and AIDS surveillance in the 1980s to monitor the immunological status of diagnosed HIV-infected individuals[179-181]. In contrast to the development of the SOPHID survey in E,W&NI, CD4 cell count monitoring was also used to monitor access to care. In addition to these case reporting systems, Scotland also introduced the voluntary reporting of all HIV testing in 1988 (including individuals testing HIV negative) and so may record time points before infection enabling the identification of seroconversion[182]. Furthermore, HIV surveillance in Scotland includes subtype monitoring and country of infection for greater characterisation of the epidemic (but until recently recorded nationality instead of ethnicity)[183]. In Australia, HIV and AIDS case surveillance is very similar to the HARS surveillance system in E,W&NI (no equivalent of CD4 Surveillance or SOPHID). Separate report forms and databases are used for HIV diagnoses and AIDS diagnoses but records can be linked between these if sufficient patient identifiers are available[184-187].

The availability of effective HAART in 1996 was the catalyst for many countries to develop HIV case reporting because back-calculation from AIDS diagnoses could no longer be used to estimate the time of infection. This issue was recognised in France but the development of confidential reporting systems delayed the introduction of HIV case reporting from 1999 until March 2003. Subsequently, virologists in France have reported all HIV diagnoses with epidemiologic and clinical information completed by a physician. Physicians only report AIDS diagnoses. The variables collected include nationality, country of residence, country of birth, occupation, reason for attending screening, previous HIV serology, route of transmission, and clinical stage but not country of infection or ethnicity. Reporting is anonymised using a unique irreversible code for each individual to maintain confidentiality but multiple reports from an individual can be linked for epidemiological monitoring. However, despite mandatory reporting and follow-up for missing epidemiologic information, under-reporting was estimated to be 30-40%[188]. Yet, France also introduced the voluntary, anonymous submission of sera from newly diagnosed individuals to test for recent infections to better inform HIV prevention strategies[189;190] (surveillance of recent infections was introduced in E,W&NI in 2008).

In the United States, HIV surveillance systems were not standardised across states and were largely limited to passive reporting until April 2008[191;192]. Until then, the variety of mandatory HIV surveillance across states included name-based reporting of all HIV-infected individuals (but allowing anonymous HIV testing), name-based reporting of symptomatic HIV-infected persons only and anonymous reporting. Other states relied on voluntary reporting systems.

Deduplication of reports was not carried out in all states with name-based reporting resulting in a significant proportion of duplicate reports[193]. The varied sources of HIV reports included state public health laboratories, private laboratories, private physicians, and other health-care providers[194;195]. HIV surveillance has been enhanced since 2007 by development of the Medical Monitoring Project, which will survey random samples of HIV-infected persons accessing care to provide detailed information of the need for care services[196].

Denmark, in the early 1990s, introduced a mandatory but anonymous HIV reporting system. Laboratory staff reported directly to the national surveillance unit but also via clinicians in order to collect patient data and to ensure maximum completeness of the reports. Serial numbers were used to link records but maintain anonymity, as long as records originated from the same laboratory or clinical source[197;198]. National HIV surveillance was very similar in Germany, with AIDS cases reported since 1982 and anonymous HIV reporting from laboratories since 1987[199] (unspecific record linkage was possible in Germany using date of birth only[176]).

Italy and Spain do not have national HIV surveillance systems despite long-running national AIDS case reporting and some regional collation of HIV reports[176]. In Italy, record linkage procedures with non-name based identifiers are used to identify newly diagnosed HIV and AIDS cases and to integrate HIV surveillance data from a variety of local sources such as hospital registers and death notifications[200-202]. In Spain, there is no possibility of record linkage due to anonymous reporting of HIV infections (only the initials were reported)[176].

The Netherlands started a new HIV reporting system in 2002 using name-based identifiers[176]. This is based on a national cohort of HIV-infected individuals founded in 2001, which was itself based upon a national cohort research study of adults receiving antiretroviral therapy undertaken from 1998 to 2000[203].

2.7 Summary

HIV case reporting systems in E,W&NI have developed over time to provide comprehensive, detailed and valid information about the HIV epidemic. The data include all diagnosed HIV infections in E,W&NI since the first identified case and are derived from reports submitted from clinics, laboratories and national death records. The completeness of the data is maximised by active follow-up for missing reports and missing information from reports. Validation and follow-up also maximise the consistency of the data. Individual-level data allow detailed presentation of data cross-tabulated by numerous different variables. Data collection balances timeliness with appropriateness such that most HARS reports are received within six months of diagnosis/death but SOPHID and CD4 reports are only expected once or twice a year.

HIV surveillance in E,W&NI is more comprehensive than surveillance in many other countries with similar HIV epidemics due to: the ability to link records from the same individual over time while maintaining confidentiality; the early monitoring of new HIV diagnoses; and the additional monitoring of prevalent diagnosed infections and levels of immunosuppression. The data are not as comprehensive as the Dutch cohort but surveillance systems have the potential to be linked to produce comparably rich information for analysis.

Chapter 3. Bilateral record linkage between three independent HIV surveillance systems

3.1 Introduction

The main methodological aim of this thesis was to develop a valid method to create a fully integrated, national HIV surveillance database by linking records from three independent HIV case reporting systems. This would allow corresponding information, which was missing from one database but not another, to be shared and allow investigation of a wider range of issues by the simultaneous analysis of data from the integrated HIV surveillance systems.

This chapter briefly describes the theory of record linkage, and then, the detailed approach taken to bilaterally link records. This approach was based on comparisons of patient-specific information between datasets and the subsequent use of an algorithm to identify reliable bilateral links between two datasets at a time (allowing for some coding errors in the data). Triangulation of these bilateral links for full integration is described in chapter 4. As HARS (HIV and AIDS Reports database) and CD4 Surveillance data were already routinely linked, this chapter focuses on the record linkage between HARS and SOPHID (Survey of Prevalent HIV Infections Diagnosed) and separately between SOPHID and CD4 Surveillance.

The record linkage algorithm was developed over time as more information was collected through SOPHID and as the SOPHID data were more completely linked over time. Therefore, the results are described in sections showing the

lessons that were learnt as the methodology developed: validation of record linkage based on soundex codes, dates of birth and sex; evaluation of record linkage including local patient ID numbers (included nationally in SOPHID reports for 2003); and record linkage between datasets using multiple records per individual.

3.2 Aims of this chapter

a) To develop a thorough understanding of the key variables used in the record linkage process.

b) To assess the use of different combinations of variables in the matching process to determine how best to link the databases allowing for some coding errors in the data to maximise sensitivity while maintaining high accuracy by only retaining reliable links between the datasets.

c) To identify factors associated with record linkage in order to consider the potential for bias to be introduced by the record linkage process, to aid the interpretation of subsequent analyses.

3.3 Introduction to the record linkage process

3.3.1 Background to record linkage

Record linkage between two databases is increasingly used, as electronic information systems proliferate, to make use of data that are already held elsewhere rather than duplicate its collection in different systems. In relation to HIV, it has been used with census data, cancer and tuberculosis (TB) registries

and vital statistics information to estimate mortality, cancer and TB-coinfection rates[204-210]. It has also been used in capture-recapture studies to estimate numbers of deaths among HIV-infected adults, cumulative numbers of HIV diagnoses, and the underreporting of AIDS cases[211-213]. If unique and reliable identifiers were available to allow every record in the datasets to be matched (often the aim of relational databases) then record linkage would be simple. However, record linkage is complicated by unreliable reporting, redundancy of information and changes over time.

Record linkage has employed a variety of methodologies depending on the aims and the information available. However, there are currently two principal conceptual strategies that are utilised. *Deterministic record linkage* looks for exact agreement between one or more (or combined) matching variables. Coding errors in the variables in one dataset are likely to result in records from the same individual not being linked. However, variables can be subdivided to find exact agreement between parts of variables, known as fuzzy matching, which can allow record linkage between variables despite coding errors (for example, the day, month and year of birth instead of the date of birth or the surname initial instead of the soundex code). In contrast, *probabilistic record linkage* tends to use a greater number of matching variables and considers the probability that two records are from the same person even if some of the variables differ, by taking into account the expected frequency of each of these variables in the dataset (i.e. rarer instances of a variable mean that it is more likely to result in record linkage if it matches with the variable from the other dataset). These methods can also be combined together or combined with

manual reviews depending on the aims of the project. Both methods involve a degree of subjectivity because cut-offs for an acceptable link between the datasets need to be decided by the user. Deterministic linkage is simple to understand, situation-specific and maintains a high positive predictive value[214]. Probabilistic linkage, on the other hand, is likely to result in greater sensitivity but usually requires prior input of variable frequencies from a reference population or an iterative process to estimate these frequencies from the study population[214-216]. Deterministic record linkage (or probabilistic linkage with careful manual review) is therefore recommended when making inferences at the individual-level based upon combined data[214;217]. Probabilistic linkage, on the other hand, tends to be used for larger population-based studies to estimate a statistical rate or prevalence[218-222].

3.3.2 Background to record linkage in HARS, SOPHID and CD4 Surveillance

The accuracy and completeness of the information reported to the HPA directly impact on the collation of data from different sources for each individual and therefore on the precision of the results that are disseminated. These are reliant on the processes of record linkage.

Record linkage within each of the three databases is referred to in this thesis as deduplication and is carried out to link reports from the same individual in order to limit over-counting and to collate information on that individual (Sections 2.2.9, 2.3.4, 2.4.4). Inability to link multiple reports of the same individual over time within each individual database (insensitive deduplication) due to incorrect

identifiers will result in an inflation of the true number of individuals and events reported. Incorrect linkage of reports from different individuals within the same database (non-specific deduplication) due to non-unique identifiers will result in the assimilation of data from different individuals and may produce inconsistencies within the dataset and lead to a reduction of the true number of individuals and events reported. Deduplication within each of the three HIV surveillance systems is deterministic although a wide array of criteria is used to allow consideration of non-unique and unreliable identifiers. Manual review of weak matches is carried out to ensure high sensitivity is not gained at the expense of specificity.

Insensitive or non-specific record linkage between datasets can have similar effects to deduplication, such as over-estimating under-reporting and producing inconsistencies by linking data from different individuals.

Record linkage between databases can link reports from the same individual that are separately held in HARS, SOPHID and CD4 Surveillance. Record linkage between HARS and CD4 Surveillance is routinely carried out to determine levels of immunosuppresion at HIV diagnosis (Section 2.3.5). Temporary matching used to be routinely carried out between SOPHID and HARS and between SOPHID and CD4 Surveillance to estimate under-reporting to SOPHID[172;223] and to determine the last date on which the individual was known to have been alive for newly diagnosed individuals[224]. Yet, the annual SOPHID datasets were not previously combined for matching purposes and

temporary matches were not used to create permanent links between the datasets. Nor were any formal analyses of the matched data undertaken.

Record linkage of SOPHID to HARS had the potential to link SOPHID and CD4 Surveillance via HARS. However, separate record linkage between SOPHID and CD4 Surveillance was undertaken because it had the potential to link a large number of CD4 Surveillance records that had not previously been linked to HARS records. This potential was conferred by the availability of local patient ID numbers requested for all biannual London SOPHID surveys and also for the national SOPHID survey since 2003. Linkage of these datasets also allowed triangulation of CD4 Surveillance, SOPHID and HARS records (Chapter 4).

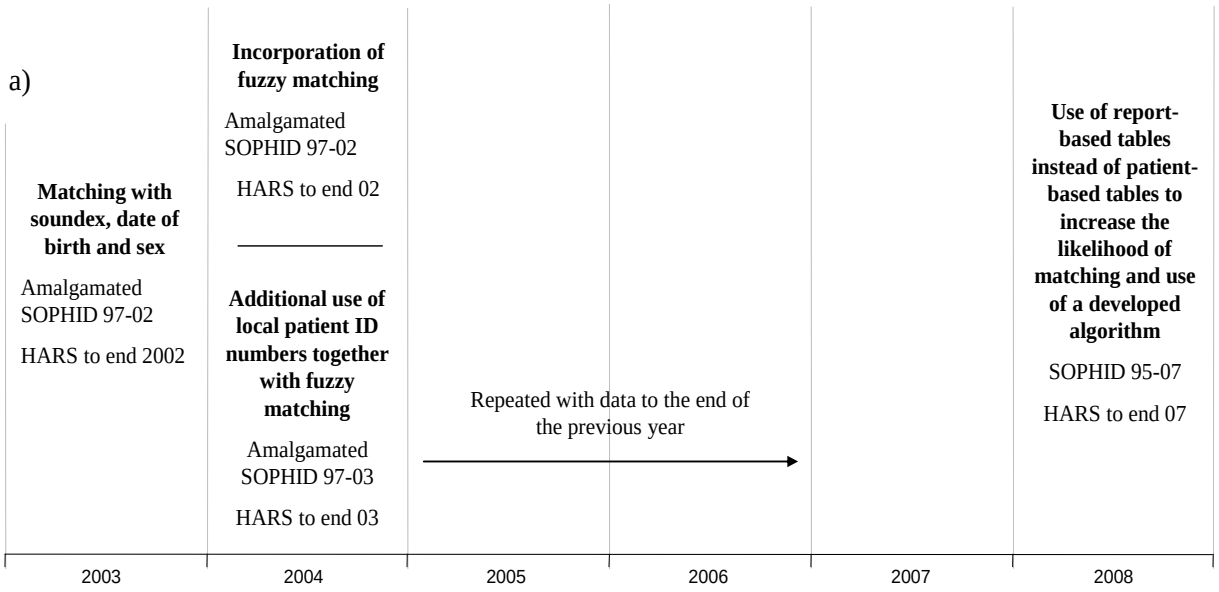
3.3.3 Development of methodology

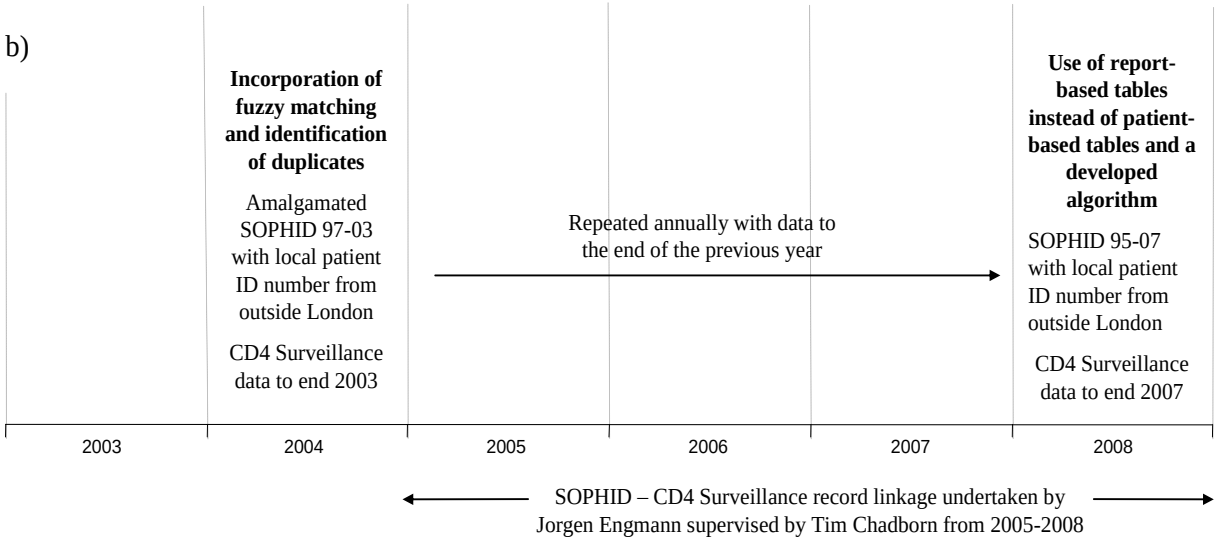
Deterministic methodology was selected as appropriate for the aims of achieving an integrated and robust dataset. This also allowed full utilisation of in-house expertise with the deterministic algorithms and software already used in the surveillance systems.

Record linkage between HARS and SOPHID originally only considered matches between patient-based tables where the soundex, date of birth and sex were exactly the same and only created unique, one-to-one links between datasets (Figure 3.1a). Fuzzy matching between HARS and SOPHID was used to improve the sensitivity of the record linkage as was incorporation of local patient ID numbers after I expanded their collection in the SOPHID surveys to areas outside London for data from 2003 (Figure 3.1a). In 2004, I devised the

SOPHID ‘Report Table’, which included all records reported since 1995, and developed the methodology for deduplicating these data and assigning a patient-specific number (Section 2.4.5). This ‘Report Table’ was then used to consider matches between multiple combinations of reported patient identifiers for the same individual and assess potential duplicates (Figure 3.1b). In 2008, methods were integrated and simplified, based on what had been learnt, and Report Tables were used for record linkage of data to the end of 2007. After this final process, triangulation between the three datasets was carried out to check the validity of multiple links to the same record and links that formed only two sides of the triangle. I carried out all of the record linkage between SOPHID and HARS and the original record linkage between SOPHID and CD4 but a colleague undertook the record linkage of SOPHID and CD4 after 2004 under my supervision (Figure 3.1a/b).

Figure 3.1 a/b. Timeline showing development of the record linkage algorithm between a) HARS and SOPHID and b) SOPHID and CD4 Surveillance





3.4 Key data used for deduplication and record linkage

3.4.1 Soundex codes

There were 5,043 different genuine soundexes reported in data to the end of December 2007 (4,467, 4,522, and 4,471 in HARS, SOPHID and CD4 datasets respectively). Within each database, about a fifth of soundexes were unique but few occurred more than 200 times (Table 3.1).

Table 3.1. Distribution of the frequency of reported soundex codes in HARS, SOPHID and CD4 Surveillance

Frequency	HARS	SOPHID	CD4 Surveillance
1	895 (20.0%)	948 (21.0%)	901 (20.2%)
2	545 (12.2%)	524 (11.6%)	501 (11.2%)
3	355 (7.9%)	332 (7.3%)	318 (7.1%)
4	254 (5.7%)	282 (6.2%)	281 (6.3%)
5	204 (4.6%)	218 (4.8%)	241 (5.4%)
6-50	1,799 (40.2%)	1,803 (39.9%)	1,798 (40.2%)
51-100	240 (5.4%)	246 (5.4%)	250 (5.6%)
101-200	121 (2.7%)	116 (2.6%)	121 (2.7%)
201-300	33 (0.7%)	34 (0.8%)	33 (0.7%)
>300	30 (0.7%)	19 (0.4%)	27 (0.6%)
Total	4,476 (100%)	4,522 (100%)	4,471 (100%)

Whereas 377 (0.5%) of 79,736 individuals reported to SOPHID and 856 (1.0%) of 87,654 individuals reported to HARS had never had a genuine soundex code reported, there were 20,225 (19.3%) of 104,619 individuals reported to CD4 Surveillance for whom all reports over time lacked soundex codes. Two-thirds (67.1% [253]) of individuals reported to SOPHID without soundex codes were

young adults (aged 15-19 years) reported from the National Study of HIV in Pregnancy and Childhood, which requests but does not require soundex codes.

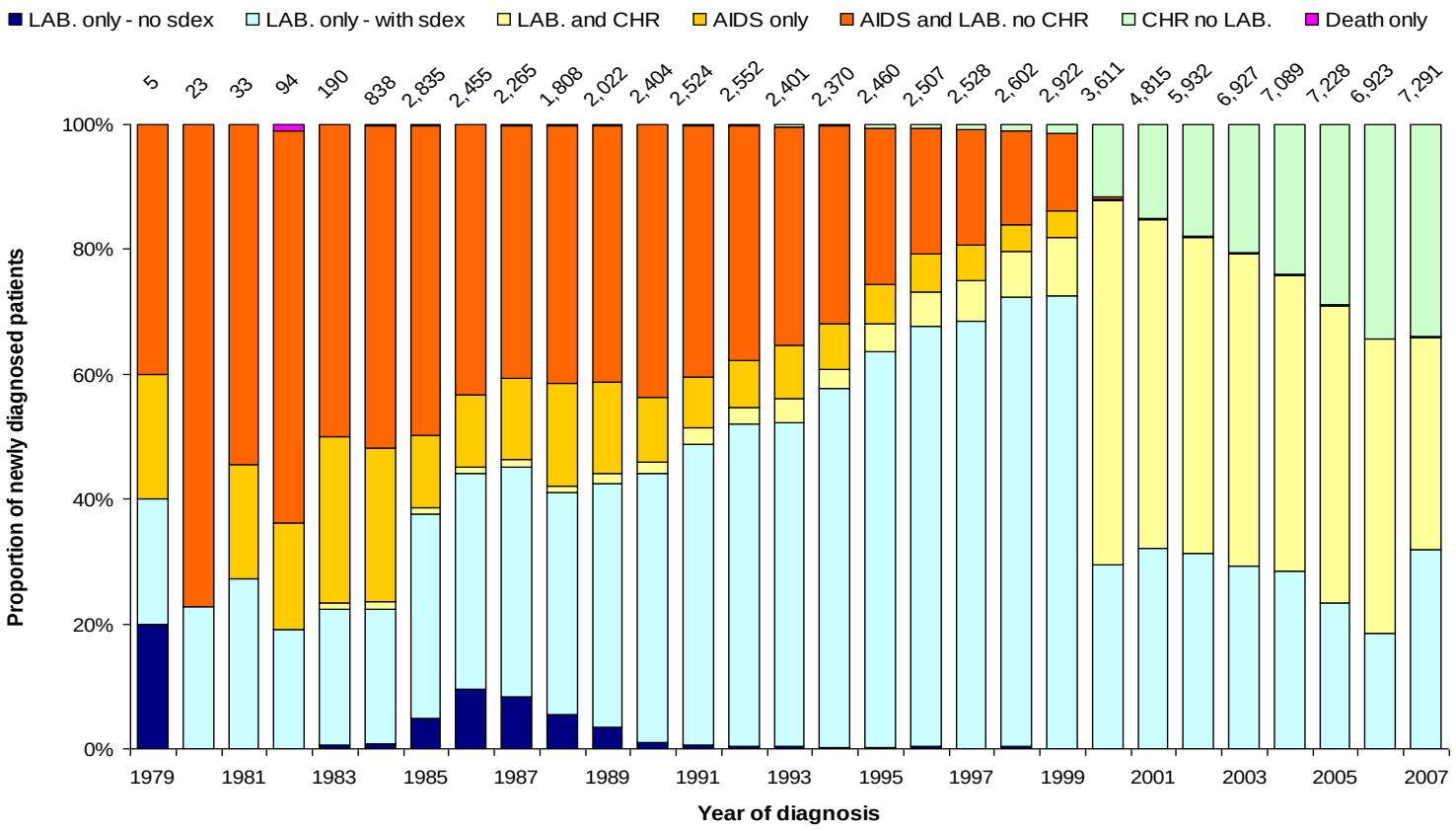
A greater number and proportion of new HIV diagnoses reported only from laboratories (822 [2.4%]) were missing soundex codes than reports from clinicians (31 [0.9%] individuals reported on an AIDS form only; 2 [0.4%] on a death form only; 1 [<0.01%] on both a laboratory form and a clinician report form; all other combinations of reports included a soundex code) (Figure 3.2). Almost all missing soundex codes were for records of new HIV diagnoses between 1985 and 1993 because only one of either soundex or local patient ID number were requested on the laboratory forms that were used at the time (Appendix B). Almost a half (47.2% [404/856]) of the individuals reported to HARS without a soundex code were reported from two London clinics[*] and, independently, almost a fifth (18.5% [158/856]) had no date of birth reported.

A high percentage of individuals with their first CD4 cell count reported to CD4 Surveillance in 2006 or 2007 were missing soundex codes (37.5% [3,534/9,433] and 40.5% [5,415/13,359] respectively compared to an average of 13.5% where the first CD4 cell count was between 1990 and 2005). These individuals accounted for 44.2% [8,949/20,225] of all individuals reported to CD4 Surveillance without soundex codes. There was no clear association between the location or size of the laboratory and the proportion of missing soundexes.

[*] However, these records only accounted for 3.7% and 4.8% of all records from these two London clinics, which were not the highest percentages from London clinics.

Figure 3.2. HARS reporting over time* by type of report# and availability of soundex code

(AIDS = AIDS report, LAB. = HIV report from laboratory)



* the first HIV infection diagnosed in the UK was in 1991 but retrospective diagnoses were made based on historical samples

10 individuals without soundex codes reported only on an AIDS report not shown due to small numbers

In each database, there were many individuals who had more than one soundex code reported: 5,308 (6.7%) of 79,359 individuals reported to SOPHID; 2,702 (3.1%) of 86,798 individuals reported to HARS; 6,493 (7.7%) of 84,469 individuals reported to CD4 Surveillance (Table 3.2). Although people may change their surnames, particularly women who marry, manual inspection showed that many individuals with multiple soundex codes had very similar codes indicating that they were likely to be due to coding errors. This was expected because records are linked due to exactly or closely matching patient identifiers. Records with substantially different soundex codes are unlikely to be linked unless they have matching dates of birth and local patient ID numbers.

Table 3.2. Distribution of the frequency of different soundex codes per individual in HARS, SOPHID and CD4 Surveillance

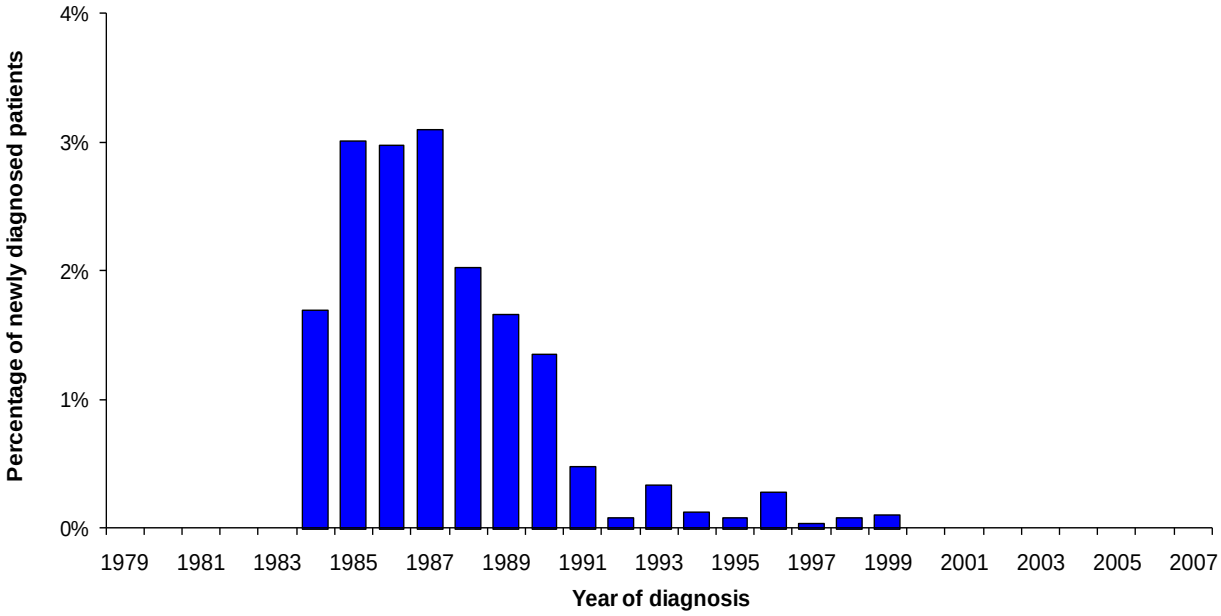
Frequency	HARS	SOPHID	CD4 Surveillance
1	84,096 (96.9%)	74,051 (93.3%)	77,976 (92.3%)
2	2,634 (3.0%)	4,878 (6.1%)	5,568 (6.6%)
3	65 (0.1%)	390 (0.5%)	751 (0.9%)
4	2 (<0.01%)	35 (0.04%)	125 (0.1%)
>4	1 (<0.01%)	5 (0.01%)	49 (0.1%)
Total	86,798	79,359	84,469

3.4.2 Dates of birth

Fewer reports were missing dates of birth than were missing soundex codes. There were 140 (0.2% of 79,736) individuals reported to SOPHID and 377 (0.4% of 87,654) individuals reported to HARS without dates of birth compared to 7,667 (7.3% of 104,619) individuals reported to CD4 Surveillance.

As with soundex codes, dates of birth were more likely to be missing in HARS for individuals diagnosed in the late 1980s (Figure 3.3). All but seven of these (370 [1.1% of all individuals who were only reported on a laboratory form]) were only reported on a laboratory form as either age or year of birth could be reported instead of date of birth on laboratory forms until 1993 (Form D93 – Appendix B.1). Two in five (39.8% [150/377]) of individuals with no date of birth were reported from three London clinics, two of which were the same London clinics that reported many individuals without soundex codes.

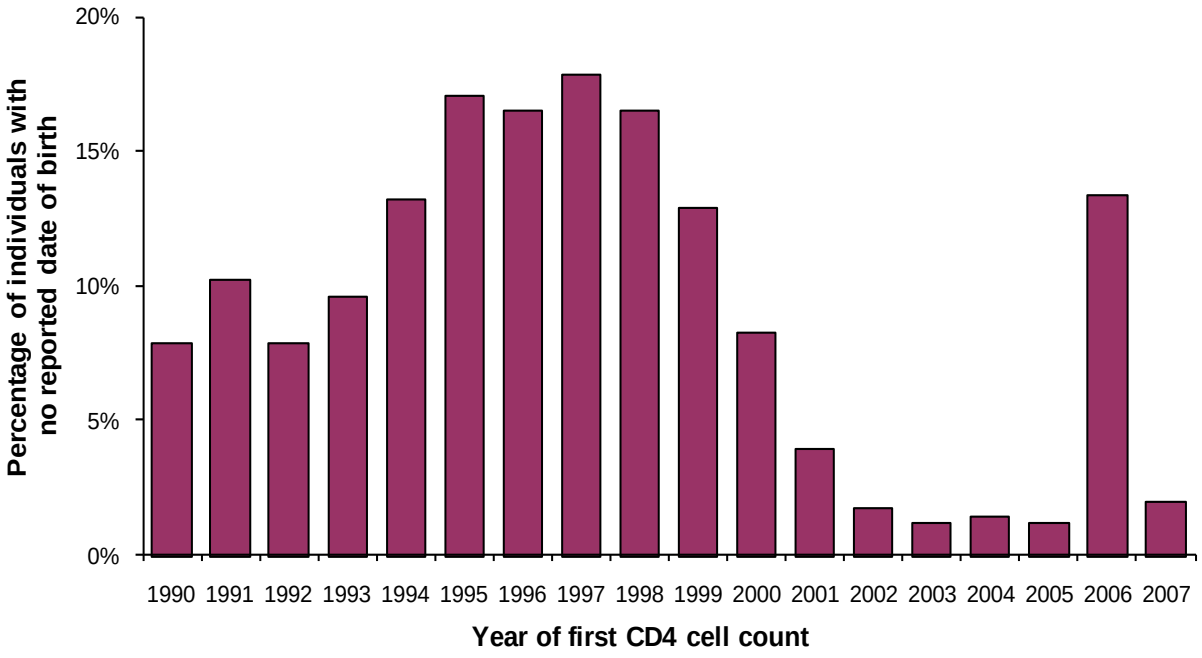
Figure 3.3. Percentage of newly diagnosed individuals without a reported date of birth.



The majority of individuals reported to SOPHID without a date of birth were only reported once between 1997 and 1999 (68.6% [96/140]). Almost a quarter (22.9% [32/140]) were reported from one of five clinics in Oxford. Only three of the 140 had no soundex code reported.

Individuals reported to CD4 Surveillance without a date of birth largely had their first CD4 cell count before 2000, although there were also a substantial number without dates of birth who were first reported in 2006 (which may have been due to changes in clinic computing systems) (Figure 3.4). There was wide variability in the reporting of dates of birth by laboratory: 75.3% of the individuals without dates of birth were reported from three large laboratories in London but there were no clear trends by location or size and there was wide variation around the average of 3.6% with missing dates of birth.

Figure 3.4. Percentage of individuals reported to CD4 Surveillance without a reported date of birth.



There were 22,460 different dates of birth reported to the end of 2007 (17,940, 16,986, and 21,316 to HARS, SOPHID and CD4 Surveillance). Within each dataset, 22-26% were unique and few occurred more than 15 times (Table 3.3).

Table 3.3. Distribution of the frequency of reported dates of birth in HARS, SOPHID and CD4 Surveillance

Frequency	HARS	SOPHID	CD4
1	3,929 (21.9%)	3,695 (21.8%)	5,635 (26.4%)
2	2,636 (14.7%)	2,512 (14.8%)	2,976 (14.0%)
3	1,965 (11.0%)	1,870 (11.0%)	2,188 (10.3%)
4	1,600 (8.9%)	1,478 (8.7%)	1,766 (8.3%)
5	1,416 (7.9%)	1,345 (7.9%)	1,499 (7.0%)
6-10	4,790 (26.7%)	4,433 (26.1%)	5,105 (23.9%)
11-15	1,372 (7.6%)	1,411 (8.3%)	1,782 (8.2%)
16-20	198 (1.1%)	204 (1.2%)	300 (1.4%)
>20	34 (0.2%)	38 (0.2%)	65 (0.3%)
Total	17,940	16,986	21,316

To examine the frequencies of reported dates of birth by individual, the date of birth in the master record of each individual in each database was considered. This was the date of birth from the earliest HIV diagnosis event reported in HARS, the first reported date of birth in CD4 Surveillance and the last reported date of birth in SOPHID. Where dates of birth were reported without a specific day of birth they were recorded as “15/xx/xxxx” if missing a day of birth and “01/01/xxxx” (some as “15/06/xxxx”, “30/06/xxxx”, or “01/07/xxxx” due to change of staff and misunderstandings) if missing both a day and month of birth. Lack of both day and month of birth was much more frequent than just a missing day of birth. In HARS, for each year of birth there was a median of 232 individuals with dates of birth on the same day and month. However, there were 855 individuals recorded as born on “01/01/xxxx”, more than twice as many as the next most frequent day and month of birth combination (401 individuals born on the “10/10/xxxx”). The “15/06/xxxx” was fifth most frequently reported (367 individuals) whereas “01/07/xxxx” and the “30/06/xxxx” were the 20th and 31st

most frequently reported combinations (289 and 281 individuals respectively). Dates associated with memorable or auspicious dates were also more frequently reported (e.g. Christmas Day)[225]. Dougan et al. suggested that the increased frequency of certain dates of birth, particularly among black Africans, reflected disproportionate duplication of records in HARS[225]. It may be more likely that these reported dates of birth are the dates of birth used consistently by separate individuals even if they are not exact dates of birth. In contrast with the conclusion of Dougan et al., I suggest that the use of these dates of birth for deduplication of records should not be compromised when used in combination with soundex codes and other patient identifiers. Similar patterns of dates of birth were evident in both the SOPHID and CD4 Surveillance databases.

In each database many individuals had more than one valid date of birth reported: 3,176 (4.0%) of 79,596 individuals reported to SOPHID; 1,984 (2.3%) of 87,277 individuals reported to HARS; 3,825 (3.9%) of 97,094 individuals reported to CD4 Surveillance (Table 3.4). Multiple dates of birth must be due to coding errors or inconsistent reporting of dates of birth by the individuals themselves (only few were also reported with missing information). For example, one individual's date of birth was reported as 21/01/19xx, 02/01/19xx and 20/01/19xx. The first two dates were reported from the same clinic with the same local patient ID number but slightly different soundex codes. The latter two dates were reported with the same soundex code but from different clinics.

Table 3.4. Distribution of the frequency of different dates of birth per individual in HARS, SOPHID and CD4 Surveillance

Frequency	HARS	SOPHID	CD4 Surveillance
1	85,293 (97.7%)	76,420 (96.0%)	93,269 (96.1%)
2	1,950 (2.2%)	3,008 (3.8%)	3,585 (3.7%)
3	32 (0.04%)	160 (0.2%)	222 (0.2%)
4	1 (<0.01%)	6 (<0.01%)	15 (0.02%)
>4	1 (<0.01%)	2 (<0.01%)	3 (<0.01%)
Total	87,277	79,596	97,094

3.4.3 Summary of key data used for record linkage and deduplication

Dates of birth were more specific to an individual than soundex codes. About a fifth of both soundex codes and dates of birth were unique, but fewer individuals were reported with the same dates of birth than with the same soundex codes. However, between one in 25 and one in 50 individuals had more than one date of birth and/or soundex code reported. There were few missing soundex codes or dates of birth in both HARS and SOPHID but a significant proportion of missing dates of birth and a substantial proportion of missing soundex codes in CD4 Surveillance. The latter is likely to have resulted in incomplete record linkage of measurements of CD4 cell counts in CD4 Surveillance and therefore gaps in patient follow-up.

3.5 Record linkage between SOPHID and HARS

3.5.1 Matching soundex, date of birth and sex: 1997 to 2002 data

This was carried out in 2003 prior to the collection of local patient ID numbers from outside London and prior to the development of the SOPHID Report Table. Therefore, only one combination of soundex, date of birth, sex, date of diagnosis/date last seen, ethnic group and risk group for each individual existed in the datasets for record linkage.

First, record linkage was carried out where soundex, date of birth and sex matched exactly. There were only a very small number of records with the same combination of soundex, date of birth and sex in each database, which had been determined to represent separate individuals. Matching of other data between HARS and SOPHID was compared to quantitatively determine whether there was evidence to support and validate this record linkage.

Second, fuzzy matching was investigated for the remainder of records where soundex, date of birth and sex did not exactly match and algorithms were developed to identify further links.

Finally, factors associated with record linkage were determined to help understand any potential biases for analyses of integrated data.

3.5.1.1 Methods

The data fields from SOPHID and HARS that were used for matching are shown below (Table 3.5).

Table 3.5. Variables in SOPHID and HARS that were used for matching.

Name (and code) of variable	**Description of data**	HIV report to HARS	AIDS report to HARS	SOPHID
Soundex (sdex)	Soundex codes consist of the first letter of the surname followed by 3 digits assigned to consonants in the surname	✓	✓	✓
Date of birth (dob)	In the form dd/mm/yyyy	✓	✓	✓
Sex	Male, female or unknown	✓	✓	✓
Ethnic group (eth)	Selected from the following options: White, black-Caribbean, black-African, black-Other, Indian/Pakistani/Bangladeshi, other/mixed	✓	✓	✓
Risk group	Selected from the following options: homosexual sex, heterosexual sex, injecting drug use, mother to child, haemophilia or blood/tissue recipient, other, unknown	✓	✓	✓
Site of care	Name of hospital or centre providing HIV-related care			✓
HIV hospital	The abbreviated name of the hospital where the patient gave the specimen for HIV testing	✓		
Laboratory	The abbreviated name of the institution or site of the laboratory that carried out the HIV test	✓		
AIDS hospital	The abbreviated name of the hospital where the patient was newly diagnosed with AIDS		✓	

Record linkage used a hierarchical algorithm of mutually exclusive and progressively less specific criteria (Table 3.6) and scores were assigned accordingly. As HARS records were linked they were excluded from further matching. Exact matches (✓) were required for the soundex, date of birth and sex variables. Less specific matches included missing (m) or discordant (X)

ethnicity, risk group or location information. Matching of locations was first carried out requiring exact matches. For locations outside London this was then repeated requiring only the first four letters of the abbreviated location name to match (L) as matching reports from outside London were more likely to relate to the same person.

Visual FoxPro 5.0 was used for the analysis.

3.5.1.2 Data preparation

There were 128,345 records relating to 43,189 individuals in the SOPHID 1997-2002 amalgamated dataset. A total of 1,872 individuals were excluded because not all of date of birth, soundex and sex were reported (19 were missing sex). Fourteen (0.03%) individuals had the same soundex, date of birth and sex combinations in the remaining dataset of 41,317 individuals but had been determined to represent separate individuals at the time of deduplication.

Individuals reported to be newly diagnosed in Scotland were excluded from an 'XLatest' extract of HARS created at the end of March 2003. Also excluded were reports of individuals who were newly diagnosed after 31 December 2002 or who were known to have died before 1 January 1997. Of the remaining 40,274 individuals, 1,034 were excluded because not all of date of birth, soundex and sex were reported. 96 (0.2%) individuals had matching soundex, date of birth and sex combinations in the remaining dataset of 39,240 individuals but had been determined to represent separate individuals.

3.5.1.3 Results

Table 3.6 shows the number of records that were linked by each individual match score, the number of records that were linked according to grouped match scores, and cumulative numbers as matches became less specific. For example, there were 1,250 links with a match score of 2 with exactly matching risk group and ethnicity and matching town/city outside London (in addition to exactly matching soundex, date of birth and sex). The locations for these records were not exact matches as these would have been linked with a match score of 1. The 713 records linked with match score 5 were less specific as information about ethnicity was missing. However, information about ethnicity was often missing in HARS and these matches were considered to be more likely than if ethnicities were discordant (match score 12).

There were 28,680 HARS records that were linked to SOPHID records as exact soundex, date of birth and sex matches (only 28,670 SOPHID records were linked to HARS records as 10 SOPHID records matched to two records in HARS with the same level of confidence) (Table 3.6). These links accounted for 69% (28,670/41,317) of records in SOPHID and 73% (28,680/39,240) of records in the HARS dataset. Ninety nine per cent of links also had matching risk group, ethnicity or location to suggest that these were indeed the same individual (match scores 1-16). In addition, 43% (12,441/28,680) of all links had complete matching information (match scores 1-2) and 79% (22,724/28,680) had at least two of ethnicity, risk group or location that matched (match scores 1-5, 7, 9-12).

Table 3.6 also shows the number of records that were linked according to grouped match scores. From the top, these links had:

- Exactly matching risk group and ethnicity (17,667; 62%)
- Exactly matching risk group but missing ethnicity (8,343; 29%)
- Exactly matching ethnicity but missing risk group (334; 1%)
- Exactly matching ethnicity and location but discordant risk group (389; 1%)
- Exactly matching risk group and location but discordant ethnicity (796; 3%).

There were 74 records in HARS that had soundex, date of birth and sex matching records in SOPHID but were not linked because they had already been linked to other records. None of these records were denoted in HARS as not being duplicates and so were flagged for future follow-up.

Of the 28,680 linked records, 93.6% (26,856) had the same risk group reported, 0.9% (245) were missing risk group information in both datasets, and a further 2.2% (645) were missing risk information in one dataset (of which, 88% (565/645) were missing in HARS). There were 933 links (3.3% of the total) with discordant risk groups (Table 3.7a).

Of the 28,680 linked records, 62.3% (17,885) had the same ethnicity reported and 2.4% (690) were missing ethnicity information in both datasets (Table 3.7b). In contrast to the risk group information, 31% (8,910) were missing information in only one dataset, of which over 99% (8,853) were missing ethnicity information in HARS. The HARS records with missing ethnicity were more likely to be white individuals such that 54% of those with known ethnicity in the linked dataset were white and 37% black African compared to 60% and 30%

respectively when SOPHID ethnicities were included. This is because missing ethnicity information is not routinely followed-up for completion of the HARS dataset. There were 1,191 links (4.2% of the total) with discordant ethnicity.

Table 3.6. Results of record linkage (✓ = exact match, L = location outside London town/city match, M = missing data in HARS, X = discordant information).

Match Score	Sdex[1]	DOB[2]	Sex	Risk group	Eth[3]	Location	Links		Grouped links	
							n	cumulative	n (%)	cumulative
1	✓	✓	✓	✓	✓	✓	11,191	11,191	17,667 (62%)	17,667 (62%)
2	✓	✓	✓	✓	✓	L	1,250	12,441		
3	✓	✓	✓	✓	✓	X	5,226	17,667		
4	✓	✓	✓	✓	M	✓	2,907	20,574	8,343 (29%)	26,010 (91%)
5	✓	✓	✓	✓	M	L	713	21,287		
6	✓	✓	✓	✓	M	X	4,723	26,010		
7	✓	✓	✓	M	✓	✓	252	26,262	334 (1%)	26,344 (92%)
8	✓	✓	✓	M	✓	X	82	26,344		
9	✓	✓	✓	X	✓	✓	351	26,695	389 (1%)	26,733 (93%)
10	✓	✓	✓	X	✓	L	38	26,733		
11	✓	✓	✓	✓	X	✓	756	27,489	796 (3%)	27,529 (96%)
12	✓	✓	✓	✓	X	L	40	27,529		
13	✓	✓	✓	X	X	✓	308	27,837	374 (1%)	27,903 (97%)
14	✓	✓	✓	X	X	L	66	27,903		
15	✓	✓	✓	X	✓	X	185	28,088	185 (1%)	28,088 (98%)
16	✓	✓	✓	✓	X	X	295	28,383	295 (1%)	28,383 (99%)
17	✓	✓	✓	X	X	X	297	28,680	297 (1%)	28,680 (100%)

[1] Soundex code

[2] Date of birth

[3] Ethnicity

Tables 3.7 a/b. Results of record linkage by a) risk group and b) ethnicity.

a)

		HARS					
		Sex between men	Injecting drug use	Blood or tissues	Heterosexual sex	Not reported/not known	Total
SOPHID	Sex between men	14,277	54	4	146	144	14,625
	Injecting drug use	77	941	1	81	16	1,116
	Blood or tissues	7	1	501	50	11	570
	Heterosexual sex	316	150	46	11,137	394	12,043
	Not reported/not known	32	13	2	34	245	326
	Total	14,709	1,159	554	11,448	810	28,680

b)

		HARS							
		White	Other/mixed	Black Caribbean	Black African	Black other	I/P/B	Not reported	Total
SOPHID	White	9,819	83	34	69	10	5	6,637	16,657
	Other/mixed	298	464	26	55	15	37	515	1,410
	Black Caribbean	17	2	557	62	12	1	192	843
	Black African	51	52	77	6,688	41	19	1,312	8,240
	Black other	29	32	34	88	126	1	120	430
	I/P/B	10	16	1	17		232	77	353
	Not reported	17	9	9	11	5	6	690	747
	Total	10,241	658	738	6,990	209	301	9,543	28,680

* I/P/B = Indian/Pakistani/Bangladeshi

Of the linked records with no reported ethnicity in HARS, 93% (8,853/9,543) had a reported ethnicity in SOPHID. Attributing the SOPHID ethnicity to those linked HARS records reduced the overall proportion with missing ethnicity in the HARS dataset from 33% (9,543/28,680) to 2% (690/28,680).

Tables 3.8 a/b. Risk group and ethnicity of individuals newly diagnosed who were linked to SOPHID a) before record linkage and b) after record linkage.

a)

		Route of infection					
		Sex between men	Injecting drug use	Blood or tissues	Heterosexual sex	Not reported/not known	Total
Ethnicity	White	7,602	590	124	1,713	212	10241
	Other/mixed	345	21	11	265	16	658
	Black Caribbean	248	11	6	454	19	738
	Black African	116	19	34	6,648	173	6990
	Black other	90	7	0	107	5	209
	I/P/B	81	4	15	191	10	301
	Not reported to HARS	6,227	507	364	2,070	375	9543
	Total	14,709	1,159	554	11,448	810	28,680

b)

		Route of infection					
		Sex between men	Injecting drug use	Blood or tissues	Heterosexual sex	Not reported/not known	Total
Ethnicity	White	12,719	1,030	435	2,335	359	16878
	Other/mixed	733	48	18	343	31	1173
	Black Caribbean	360	16	8	518	28	930
	Black African	173	24	51	7,753	301	8302
	Black other	167	8	0	142	12	329
	I/P/B	107	4	26	230	11	378
	Not in HARS or SOPHID	450	29	16	127	68	690
	Total	14,709	1,159	554	11,448	810	28,680

* I/P/B = Indian/Pakistani/Bangladeshi

When the 10,560 HARS records that did not link to SOPHID were included in the assessment of missing ethnicity information, SOPHID data reduced the proportion with missing ethnicity in HARS from 42% (16,314/39,240) to 19% (7,461/39,240). The percentage missing ethnicity information by risk group was reduced from: sex between men, 50% to 20%; heterosexual sex, 25% to 12%; injecting drug use, 60% to 37%; blood transfusions/blood or tissue transfer, 67% to 20%; not known, 62% to 41%.

3.5.1.4 Investigation of the subset not yet linked at one site

A 'small' site with 62 SOPHID records and 53 HARS records was selected for further investigation of records that had not yet been linked. The aim was to determine whether HARS records that had not yet been linked were also reported to SOPHID but with a different soundex, date of birth or sex. There were 13 HARS records that were not linked to SOPHID records on the basis of exact soundex, date of birth and sex. Records reported to SOPHID were trawled for records matching HARS records on either the date of birth or soundex. Six of these records could be identified in SOPHID with a high degree of confidence (Table 3.9). This indicated that the maximum sensitivity of record linkage using exactly matching soundex, date of birth and sex was only 87% (40/46). Further matches that were not identified, due to either different soundex and date of birth reported or because patients were reported from elsewhere, would reduce this sensitivity.

Table 3.9. Matches at one 'small' site between HARS and SOPHID without all of soundex, date of birth and sex matching (✓ = exact match)

Number matched	Soundex	Date of birth	Sex	Risk group	Ethnicity	Location
1	✓	✓		✓	✓	✓
1	First letter	✓	✓	✓	✓	✓
1	Numbers correct Letter wrong	✓	✓	✓	✓	✓
2	✓	1 digit incorrect in day	✓	✓	✓	✓
1	✓	1 digit incorrect in day	✓	✓	Missing in SOPHID	✓

This brief investigation showed the potential for fuzzy matching and therefore algorithms were developed to identify further links for the remainder of the HARS and SOPHID records that had not previously been linked.

3.5.1.5 Record linkage of the unlinked with fuzzy matches

Fuzzy matching was carried out in a similar manner to the hierarchical record linkage based on matching soundex, date of birth and sex described above. However, for the first four match scores, discordant sex was allowed for records with exactly matching soundex and date of birth (Table 3.10). Subsequently, records were linked with only either date of birth or soundex matching exactly. The other of these patient identifiers was allowed to differ slightly between linked records. For these fuzzy matches, additional matching information was considered essential to limit false links. As before, less specific links with other information missing or discordant were allowed as the match score increased.

As dates of birth have less redundancy than soundex codes in these databases (Section 3.4.3) matches were sought with the same date of birth prior to matches with the same soundex. Due to the redundancy of soundex codes, soundex matches required a fuzzy date of birth match whereas some date of birth matches were allowed with completely discordant soundex if there was sufficient supporting information. Some manual checking of individual records was carried out to limit false links.

A total of 1,925 HARS records were linked to SOPHID records using fuzzy matching. The majority (1,139; 59%) of these fuzzy matches had exactly

matching soundex and sex but slightly different dates of birth. Most of the records linked with fuzzy date of births had two parts of the date of birth matching but a different day, month or year. Only 5% of the records linked with fuzzy matching had exact soundex and date of birth but different sex.

Fuzzy matches linked 18.2% of the 10,560 previously unlinked HARS records (39,240 records with soundex and date of birth minus 28,680 linked records) and accounted for 6.3% (1,925/30,605) of the total number of links made between the datasets. However, fuzzy matching only increased the proportion of linked records in HARS from 73% (28,680/39,240) to 78% (30,605/39,240).

Of the 1,925 fuzzy matches, 88.9% (1,712) had the same risk group reported, which was a similar proportion to that for the exact soundex, date of birth and sex matches. There were 98 links (5.1% of the total) with discordant risk groups (compared to 3.3% for exact matches), 11 (0.6%) with missing risk group information in both datasets, and 104 (5.4%) missing information in one dataset.

Similar proportions of the 1,925 fuzzy matches as exact matches had matching ethnicity: 56.3% (1,083) had the same ethnicity; 3.0% (57) had missing ethnicity information in both datasets; 3.5% (68) had discordant information; and 37.2% (717) were missing information in only one dataset, of which 96.0% (688) were missing ethnicity in HARS. Of the newly linked records with no reported ethnicity in HARS, 92.3% (688/745) had a reported ethnicity in SOPHID.

Table 3.10. Results of fuzzy matching (✓ = exact match, L = location outside London town/city match, X = discordant information, empty cell = not considered).

Match score	Soundex	Date of birth	Sex	Risk group	Ethnicity	Location	Links	Cumulative totals	Grouped totals (%)
1	✓	✓	X	✓	✓		43	43	90 (5%)
2	✓	✓	X	✓	X		18	61	
3	✓	✓	X	X	✓		15	76	
4	✓	✓	X	X	X	✓	14	90	
5	First letter and following two numbers	✓	✓	✓	✓		139	229	696 (36%)
6	First letter	✓	✓	✓	✓	✓ or L	254	483	
7	Three numbers only	✓	✓	✓	✓	✓ or L	80	563	
8	X	✓	✓	✓	✓	✓ or L	30	593	
9	First letter and following number	✓	✓	✓	missing	✓ or L	69	662	
10	First letter and following two numbers	✓	✓	✓	missing		57	719	
11	First letter and following number	✓	✓	✓	missing	X	67	786	
12	✓	Americanised date of birth	✓	✓	✓		32	818	1,139 (59%)
13	✓		✓	✓	X		23	841	
14	✓		✓	X	X		10	851	
15	✓	Switched day (i.e. 12 vs 21)	✓			✓ or L	19	870	
16	✓	Switched month (i.e. 12 vs 21)	✓			✓ or L	19	889	
17	✓	Switched year (i.e. 12 vs 21)	✓			✓ or L	8	897	
18	✓	Discordant month of birth	✓	✓	✓		99	996	
19	✓		✓	✓	X		107	1,103	
20	✓		✓	X	X	✓ or L	35	1,138	
21	✓	Discordant day of birth	✓	✓	✓		189	1,327	
22	✓		✓	✓	X		171	1,498	
23	✓		✓	X	X	✓ or L	57	1,555	
24	✓	Discordant year of birth	✓	✓	✓		156	1,711	
25	✓		✓	✓	X		149	1,860	
26	✓		✓	X	X	✓ or L	65	1,925	

Of all 30,605 linked records, 11 (0.04%) HARS records had dates of death prior to being first reported to SOPHID, of which 5 were fuzzy matches. Five of the 6 records linked with exact soundex, date of birth and sex were also reported from matching clinics and all were recorded to have died in 1997. This suggests that errors occurred in record linkage between HARS and ONS records in 1997, which was the first year that ONS quarterly files were used. The five fuzzy matches all had different dates of birth between HARS and SOPHID and were reported from different clinics suggesting incorrect record linkage.

Additionally, 95 (0.3%) individuals were recorded in HARS as having died in a calendar year prior to being last reported to SOPHID, of which 11 were fuzzy matches. Later investigation of some of the individuals last reported to SOPHID in 2002 indicated that they continued to be seen for care after 2002 suggesting that most of these HARS dates of death were likely to be incorrect.

In addition, of the linked records, 521 (1.7%) records were in HARS as newly diagnosed in a calendar year after being first reported to SOPHID, of which 62 (12%) were fuzzy matches. There was a mean of 2.0 years between the year first reported to SOPHID (1997-2002 dataset) and the year first reported to HARS. Although it is clear that some of these differences may be due to incorrect record linkage, it is likely that the majority reflect errors in HARS due to failure to deduplicate HARS records sufficiently (linking the matched HARS record to an earlier HARS report), coding errors or delayed reporting with incorrect dates of HIV diagnosis. However, there were only 41 of cases where

the first report received was from a clinician at the time of an AIDS diagnosis and there was no previous report of a new HIV diagnosis from the laboratory.

3.5.1.6 Factors associated with record linkage

The third aim of this chapter was to identify factors associated with record linkage in order to consider the potential for bias to be introduced into analyses of the integrated data by the record linkage process. Therefore the associations between epidemiologic factors and record linkage were considered by multivariable logistic regression analysis with successful and unsuccessful record linkage as the binary outcome (Table 3.11). All of the following factors were included in the adjusted model. This showed the following:

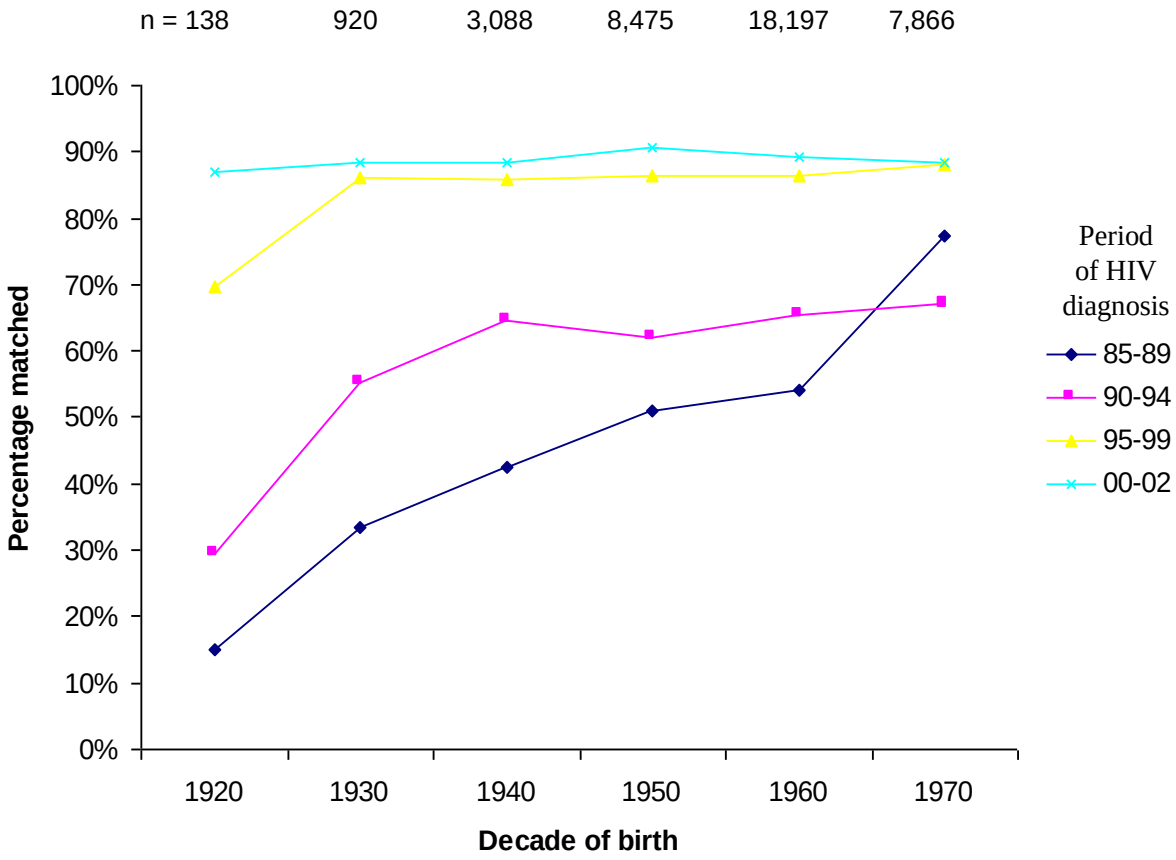
- The proportion linked increased with increasing period of HIV diagnosis
- Fewer individuals aged over 39 years were linked compared to those aged 30-34 years. Although numbers were small, a higher proportion of patients aged younger than 15 years were linked than for other age groups.
- IDU and heterosexuals were about half as likely to be linked as MSM. Individuals infected through 'other/not known' transmission were less likely to be linked (adjusted odds ratio [AOR] 0.17; 95% CI 0.15, 0.20).
- Slightly more individuals diagnosed outside London were linked compared to individuals diagnosed in London (AOR 1.16; 95% CI 1.10, 1.23).
- White individuals were more likely than individuals of other ethnicities to be linked. A lower proportion of individuals of unknown ethnicity were linked than individuals of white ethnicity (AOR 0.29; 95% CI 0.27, 0.31).

Table 3.11. Number and percentage of HARS records linked to SOPHID

	Number of linked HARS patients n (distribution, %)	All newly diagnosed patients n (distribution, %)	Percentage of newly diagnosed patients linked %	Multivariable adjusted odds ratio (95% confidence interval)
Period of HIV diagnosis				
1979-1984	285 (0.9)	378 (1)	**75**	**0.18 (0.14, 0.24)**
1985-1989	2,533 (8)	4,999 (13)	**51**	**0.12 (0.11, 0.13)**
1990-1994	5,274 (17)	8,228 (21)	**64**	**0.24 (0.22, 0.27)**
1995-1999	10,670 (35)	12,308 (31)	**87**	**0.77 (0.71, 0.84)**
2000-2002	11,843 (39)	13,327 (34)	89	-
Age group at diagnosis				
<15	154 (0.5)	170 (0.4)	**90**	**4.21 (2.40, 7.38)**
15-24	4,063 (13)	5,497 (14)	74	1.07 (0.98, 1.17)
25-29	7,050 (23)	9,236 (24)	76	1.04 (0.96, 1.12)
30-34	7,495 (24)	9,454 (24)	79	-
35-39	5,394 (18)	6,757 (17)	80	0.93 (0.85, 1.01)
40-44	2,862 (9)	3,624 (9)	**79**	**0.88 (0.80, 0.98)**
45+	3,587 (12)	4,502 (11)	**80**	**0.82 (0.74, 0.90)**
Risk group				
MSM	15,517 (51)	19,585 (50)	79	-
IDU	1,257 (4)	2,080 (5)	**60**	**0.46 (0.41, 0.51)**
Blood/blood products	598 (2)	743 (2)	**80**	**1.91 (1.53, 2.40)**
Heterosexual men	4,880 (16)	6,170 (16)	**79**	**0.51 (0.47, 0.55)**
Heterosexual women	7,496 (24)	9,140 (23)	**82**	**0.58 (0.54, 0.63)**
Other/not known	857 (3)	1,522 (4)	**56**	**0.17 (0.15, 0.20)**
Region of diagnosis				
London	19,053 (62)	24,950 (64)	76	-
Outside London	11,552 (38)	14,290 (36)	**81**	**1.16 (1.10, 1.23)**
Ethnicity				
White	10,741 (35)	11,988 (31)	90	-
Black African	7,565 (25)	8,622 (22)	**88**	**0.71 (0.64, 0.79)**
Black Caribbean	782 (3)	887 (2)	**88**	**0.71 (0.57, 0.88)**
Black Other	210 (0.7)	242 (0.6)	**87**	**0.62 (0.42, 0.91)**
Indian\Pakistani\Bangladeshi	322 (1)	389 (1)	**83**	**0.57 (0.43, 0.76)**
Other/mixed	697 (2)	807 (2)	**86**	**0.63 (0.50, 0.78)**
Not known	10,288 (34)	16,305 (42)	**63**	**0.29 (0.27, 0.31)**
Total	30,605 (100)	39,240 (100)	78	-

In addition, the proportion linked was determined for various patient identifiers (Appendix A.14). Analysis showed that decade of birth was a factor associated with record linkage but confounded by the period of diagnosis. Older individuals diagnosed between 1985 and 1989 were less likely to be reported to SOPHID. This trend became less evident as the period of diagnosis became more recent suggesting that it may have been related to survival (Figure 3.5).

Figure 3.5. Percentage of HARS records linked to SOPHID records by decade of birth for different periods of diagnosis (1985 to 2002).



The site (HIV hospital) that reported the HARS record was also considered as a factor associated with record linkage. For individual sites, the range in the percentage linked was 15% to 100%. Reports were grouped together according to the total number of reports received from each site to investigate whether there were trends by size of site. The percentage linked increased with the number of patients reported except for the largest sites (Table 3.12).

The percentage of reports linked was also considered as a function of the SOPHID reporting site. For individual sites the proportion linked ranged from 7% to 99%. A similar but more marked trend in the percentage linked was seen when grouped by the number of patients reported than for HARS sites (Table 3.12). As for HARS sites, the percentage linked was low for the largest sites.

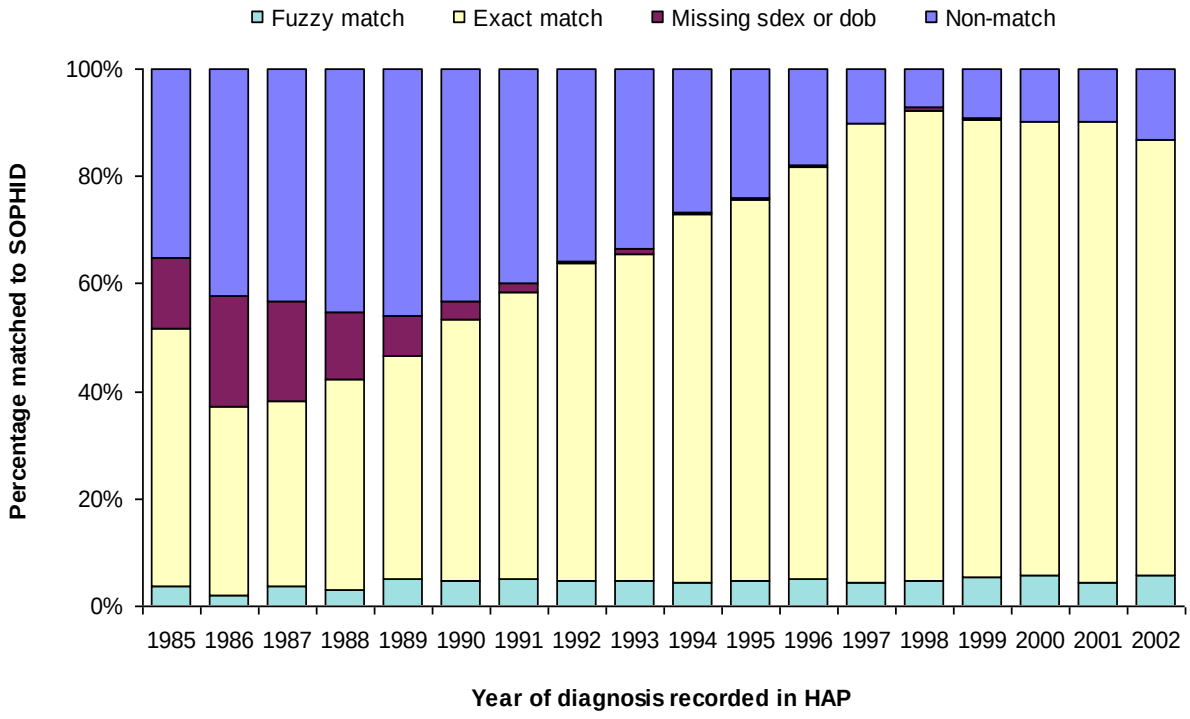
Table 3.12. Percentage of records linked by the total number of reports per site

	HARS records linked to SOPHID		**SOPHID records linked to HARS**	
Total number of reports	**Number of reporting sites**	**Percentage of patients linked**	**Number of reporting sites**	**Percentage of patients linked**
1-9	294	72%	85	53%
10-24	112	73%	52	59%
25-49	66	75%	58	66%
50-74	33	77%	29	71%
75-99	16	77%	22	71%
100-149	22	76%	25	74%
150-199	15	81%	11	75%
200-499	22	80%	31	78%
500+	10	75%	15	73%

3.5.1.7 HARS reports with missing soundex, date of birth or sex

The 1,034 records in HARS with missing soundex, date of birth or sex were added to the 8,635 HARS records that did not link and the 30,605 records that did link to examine the proportion of all HARS records linked to SOPHID records (Figure 3.6). A higher proportion (92% in 1998) of individuals recently diagnosed was linked to SOPHID than those diagnosed earlier (42% in 1988). Furthermore, the 1,034 individuals missing soundex, date of birth, or sex were mostly diagnosed in the mid-1980s and could not be linked to SOPHID records to confirm their survival. Many of these individuals may have died but even if their deaths were reported to HARS, it will not have been possible to link these events to their diagnoses due to the missing patient information.

Figure 3.6. Percentage of HARS records linked to SOPHID records by year of diagnosis (1985 to 2002).



3.5.1.8 Discussion

Exact soundex and date of birth matches linked almost three-quarters of records in the HARS and SOPHID datasets that could be potentially matched. Almost all the linked records with exactly matching soundex and date of birth also had additional matching information to suggest that these were indeed the same individual. In addition, almost a fifth of the HARS records that did not match on soundex, date of birth and sex were linked with good confidence using fuzzy matches. There was little discordance of sex, risk group, ethnicity or dates of diagnosis/death between linked records but many records in the HARS dataset were missing ethnicity information (Section 3.5.1.3). A benefit of record linkage is that much of the missing ethnicity information can be gathered from SOPHID. Discordant risk groups were expected because information collected through confidential follow-up for HARS may differ from information in clinic records. However, it is likely that much of the discordant information (5% or fewer of linked records) was due to clerical errors in the data collection or data transcription process. The discordance could also be explained as false links between datasets or false deduplication but it is likely that the use of exact soundex, date of birth and sex, which are duplicated few times within a dataset, conferred a high specificity. The low number of matching HARS and SOPHID records that were not linked because one of the records had already been linked provides supporting evidence for this high specificity. For example, it is unlikely that an error in the date of birth would lead to a match with another individual with the same soundex. However, discordant dates of death suggested that some false links were created by fuzzy matching.

Although the matching process was limited to individuals with complete soundex, date of birth and sex information who were expected to feature in both datasets, over a fifth of HARS records could not be linked to SOPHID. Reasons for this may be that individuals with HARS records had died without our knowledge or emigrated prior to 1997, individuals had changed names since their original diagnosis or provided false information to maintain anonymity, that individuals were reported to HARS but not to SOPHID (under-reporting to SOPHID), or that patient identifiers were reported differently and so could not be linked. All of these factors may account for part of the large number of unmatched records but are considered in more detail below:

- Most deaths should have been determined through matching records to ONS records if they were not reported directly from physicians. However, over 1,000 individuals mostly diagnosed in the mid-1980s had insufficient identifiers reported to confirm their death or survival;
- Relatively few individuals were expected to have changed their names between HIV diagnosis and inclusion in SOPHID surveys;
- The number of individuals who provided false information to maintain anonymity is likely to have been minimised because of the confidential service that GUM clinics provide[226]. False names may have been more frequently used in the 1980s and early 1990s when there was more fear about disclosure of a terminal disease, in particular, due to restrictions for life insurance[227;228]. However, patient and NHS websites advise people of the confidentiality of GUM services and of the option of using a false name but emphasise the need to use that same false name again when

reattending. I could not find any similar advice about the use of false dates of birth. If individuals continued to use the same false information, record linkage would still have been effective but this consistency may be less likely if patients change clinics;

- If individuals diagnosed in England and Wales had emigrated before 1997 they would have appeared in HARS but not in SOPHID. However, a similar number of SOPHID and HARS records were considered for record linkage. If many individuals had emigrated, substantial under-reporting to HARS or over-counting in SOPHID must have balanced these numbers;
- A substantial number of individuals reported to HARS but not to SOPHID was unlikely because a similar number of SOPHID records were not linked to HARS. It was unlikely that there was such a degree of under-reporting to both surveillance systems, in particular, because SOPHID data are used to allocate funding on the basis of the number of individuals reported;
- Coding errors that became evident through fuzzy matching accounted for a fifth of the links that did not link on exact soundex, date of birth and sex. Coding errors were also likely to account for most of the discordant risk groups and ethnicity from linked records. It is likely that many other links could not be identified through fuzzy matching.

3.5.2 Use of local patient ID numbers in the SOPHID 1997-2003 dataset for record linkage to HARS data (2004)

3.5.2.1 Data preparation

Data preparation for record linkage in 2004 between SOPHID and HARS was the same as in 2003 (Section 3.5.1.2) except that local patient ID numbers were available for the majority of individuals reported to the 2003 SOPHID survey and included in the amalgamated dataset. Of the 34,260 records in the deduplicated SOPHID dataset for 2003 with all of date of birth, soundex and sex reported, 99.8% (19,434/19,473) of records from the London survey and 88.5% (13,093/14,787) from outside London had local patient ID numbers reported.

The dataset from the 2003 SOPHID survey was appended to the previous amalgamated dataset with the additional local patient ID numbers available for matching and amalgamated as previously. There were 158,459 records relating to 49,197 individuals in the SOPHID 1997-2003 dataset. Records for 107 individuals were excluded because not all of date of birth, soundex and sex were reported to leave a dataset of 49,090 individuals. There were 32,484 (66.2%) records with a local patient ID number of which 29,249 (90.0%) were longer than four characters (to provide at least some additional specificity for records containing ‘GUM.....’ or similar at the start of their local patient ID number [Section 2.1.2]) and could be used for reliable matching.

The HARS dataset from the end of June 2004 included records for 47,349 individuals, 1,027 were excluded for whom not all of date of birth, soundex and sex were reported to leave a dataset of 46,322 individuals.

3.5.2.2 Record linkage process

Local patient ID numbers were included in this process as it was found to be very effective in record linkage between SOPHID and CD4 Surveillance. Reports in the SOPHID and HARS databases were linked as previously using exact and fuzzy matched patient information, but including local patient ID numbers. However, the process was streamlined because the results of the previous record linkage strategy suggested that many records could be linked with high confidence. For example, all records with matching soundex code, date of birth and sex were linked without considering risk group, ethnicity or location. Matches were not considered for record linkage if they had a date of death in HARS prior to the year that they were first reported to SOPHID. Visual FoxPro 5.0 was used for the analysis.

3.5.2.3 Results

There were 38,975 records linked between the SOPHID and HARS datasets (Table 3.13). These accounted for 79.4% of the 49,090 SOPHID records and 84.1% of the 46,322 HARS records. Of the total, 34,783 (89.2%) were linked by exact soundex, date of birth and sex, 1,360 (3.5%) were linked using the local patient ID number but did not have exact soundex, date of birth and sex matching, 701 (1.8%) were linked by exact date of birth and sex but fuzzy soundex and 2,131 (5.5%) by exact soundex and sex but fuzzy date of birth.

Nine per cent (1,360) of the 15,045 patient ID number matches did not have matching soundex, date of birth and sex but some of these could have been linked with fuzzy soundex and date of birth. Local patient ID numbers only

corroborated 39.1% (13,685) of the 34,973 exact soundex, date of birth and sex matches and only half (15,045/29,249) of the SOPHID records with a sufficiently long patient ID number were linked. It is important to note here that the patient ID number from the clinic at the time of HIV diagnosis was being matched to the patient ID number from the clinic of last attendance in 2003.

There were eight links where an individual was reported to SOPHID in a calendar year after the year of death in HARS. These were investigated individually and all dates of death were subsequently believed to be incorrect. There were 32 links where the individual was reported to SOPHID as having died and the last year reported to SOPHID was earlier than the year of death reported to HARS. However, these HARS deaths were all reported to be in January or February of the year after they were last reported to SOPHID and the links were therefore believed to be reliable. There were also three links where the year of death reported to SOPHID was earlier than the year of diagnosis reported to HARS and all of these linked incorrectly with a fuzzy date of birth. There were 60 additional links where the last year of report to SOPHID (but no death reported) was earlier than the year of diagnosis reported to HARS. Half (32) of these were linked incorrectly with a fuzzy date of birth (31) or fuzzy soundex code (1), a different reporting clinic and no matching local patient ID number. The other 28 were believed to be correct links, mostly accounted for by HARS reports that were not linked to previous HARS reports, but where SOPHID reports had been linked over time (through deduplication). However seven of these 28 appeared to be because diagnoses in early January were reported to SOPHID as seen at the end of the previous calendar year.

3.5.2.4 Discussion

Local patient ID numbers provided a highly specific match with only additional soundex or date of birth matching needed to ensure that records with similar patient ID numbers at different clinics/centres were not linked. However, although patient ID numbers provided confirmation of true links with discordant soundex or date of birth, they did not substantially increase the number of records linked between HARS and SOPHID datasets. This is likely to be because local patient ID numbers are specific to a clinic and the majority of reports from a clinic are likely to have matching soundex codes and dates of birth. Links based on soundex code and date of birth but with different local patient ID numbers were most likely to be due to patients changing clinic between diagnosis and when last seen for care. Links between SOPHID and HARS records were not reliable if the year of death reported to SOPHID was earlier than the year of diagnosis reported to HARS. However, if the individual was reported to SOPHID after the date of death recorded in HARS, the link was reliable and the date of death was not. Although the numbers of probable incorrect links (found through analysis of discordant dates) was small, all of these were based on fuzzy matches, indicating that the additional sensitivity gained is not worth the probable loss of specificity.

Table 3.13. Results of record linkage (✓ = exact match, X = discordant information, L = town/city match outside London, missing = ethnicity missing in HARS dataset, empty boxes not considered).

Match Score	Sdex[1]	Dob[2]	Sex	Risk	Eth[3]	Loc[4]	Patient number	Individual links		Grouped links	
								n	cumulative	n (%)	cumulative
1	✓	✓	✓				✓	11,060	11,060	34,783 (89%)	34,783 (89%)
2	✓	✓	✓				Fuzzy	2,625	13,685		
3	✓	✓	✓				X	21,098	34,783		
4	✓	✓	X				✓	41	34,824	1,360 (3%)	36,143 (93%)
5	X	✓					✓	516	35,340		
6	✓	X					✓	353	35,693		
7	X	X				✓/L	✓	62	35,755		
8		✓					Fuzzy	248	36,003		
9	✓						Fuzzy	140	36,143		
10	✓	✓						130	36,273	701 (2%)	36,844 (95%)
11	1,3[5]	✓	✓	✓	✓			111	36,384		
12	1,1[5]	✓	✓	✓	✓	✓/L		190	36,574		
13	2,3[5]	✓	✓	✓	✓	✓/L		87	36,661		
14		✓	✓	✓	✓	✓/L		26	36,687		
15	1,2[5]	✓	✓	✓	missing	✓/L		79	36,766		
16	1,3[5]	✓	✓	✓	missing			78	36,844		
17	✓	Am dob[6]	✓					77	36,921	2,131 (5%)	38,975 (100%)
18	✓	Flip dD[6]	✓			✓/L		25	36,946		
19	✓	Flip mM[6]	✓			✓/L		20	36,966		
20	✓	Flip yY[6]	✓			✓/L		0	36,966		
21	✓	Dd/--/Yy[6]	✓	✓	✓			151	37,117		
22	✓	Dd/--/Yy[6]	✓	✓	X			263	37,380		
23	✓	Dd/--/Yy[6]	✓	X	✓			36	37,416		
24	✓	--/Mm/Yy[6]	✓	✓	✓			273	37,689		
25	✓	--/Mm/Yy[6]	✓	✓	X			480	38,169		
26	✓	--/Mm/Yy[6]	✓	X	✓			54	38,223		
27	✓	Dd/Mm/--[6]	✓	✓	✓			238	38,461		
28	✓	Dd/Mm/--[6]	✓	✓	X			463	38,924		
29	✓	Dd/Mm/--[6]	✓	X	✓			51	38,975		

[1] Soundex, [2] Date of birth, [3] Ethnicity, [4]Location

[5] Substring of soundex code: 1,3 = first letter and following two numbers, 1,2 = first letter and following number, 1,1 = first letter, 2,3 = second number and following two numbers

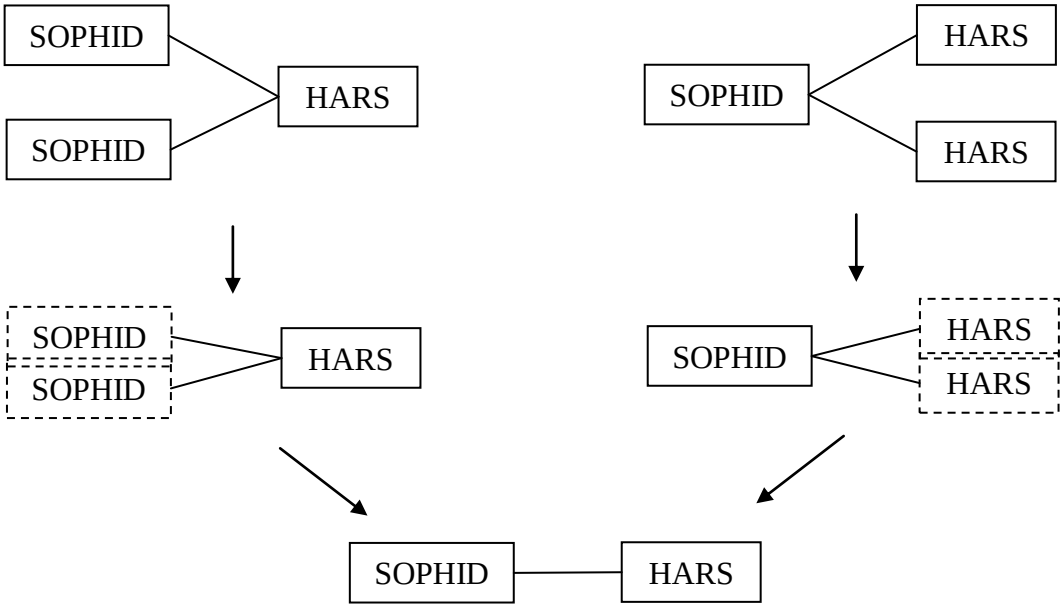
[6] Fuzzy date of birth: Am dob = Americanised date of birth, flip = switched number (i.e. 12 vs 21), Dd/--/Yy = discordant month of birth, --/Mm/Yy = discordant day of birth, Dd/Mm/-- = discordant year of birth

3.5.3 Record linkage between HARS and SOPHID using multiple records for each individual: 2008

While the central methodology of record linkage between HARS and SOPHID remained the same in 2008, the datasets were modified. These datasets included all reports to HARS and SOPHID (Report Tables) rather than deduplicated data with records unique to each patient. The aim was to maximise the record linkage between the databases by enabling matching between multiple reports of an individual rather than only between unique aggregated records for each individual. When using an aggregated record for an individual, a single soundex, date of birth and local patient ID number must be chosen to represent the patient. However, using all reports for a patient meant that all different combinations of the patient identifiers were considered for record linkage.

A key difference to the record linkage strategy in 2008 was that the algorithm was not mutually exclusive and records were not excluded from subsequent matching. Therefore, records could be linked a number of times between the datasets. This identified probable duplicates and therefore, an additional step of deduplicating records in the HARS and SOPHID matching tables was incorporated in the process. This resulted in unique one-to-one links between the final datasets.

Figure 3.7. Schema summarising how bilateral links between HARS and SOPHID were deduplicated



3.5.3.1 Data preparation

There were 562,130 records relating to 80,008 adults in the SOPHID 1995-2007 dataset. Individuals aged younger than 15 years were excluded as numbers were small and because many did not have a soundex code reported. Records were not excluded because of missing date of birth, soundex or sex as these records could still be linked despite missing information. There were no individuals with matching patient identifiers in the final SOPHID dataset as these were all deduplicated.

The HARS reports used for record linkage included data from a combined extract of the XLatest and Report Tables at the end of June 2008. For every individual, this incorporated every unique combination as reported. Reports of individuals who were diagnosed when aged less than 15 years, who were newly

diagnosed after 31 December 2007 or who were known to have died before 1 January 1995 were excluded. There were reports from 79,192 individuals among the remaining 130,492 records, of whom 728 individuals did not have all of date of birth, soundex and sex reported. There were 59 (0.08%) individuals with matching soundex, date of birth and sex combinations in the remaining dataset, of which 11 had matching ethnicities and risk groups.

3.5.3.2 Record linkage process

Reports in the SOPHID and HARS databases were linked as previously using exact and fuzzy matched patient information (Section 3.5.2.2). However, the process was streamlined even further so that only robust links were included in line with the SOPHID deduplication algorithm of the Report Table (Section 2.4.6). However, this strategy did not incorporate fuzzy soundex code and date of birth matching as the additional yield was not considered worth the possible loss of specificity without manual inspection of these potential links. As previously, matches were not considered if they had a date of death in HARS prior to the year that they were first reported to SOPHID. Matches were not considered if the SOPHID year of death was earlier than the year of HIV diagnosis reported to HARS (Section 3.5.2.4).

As noted above, this process was not mutually exclusive and records were not excluded from subsequent matching, which therefore required additional deduplication in the HARS and SOPHID datasets.

Microsoft Access 2003 was used for the analysis.

3.5.3.3 Results

After record linkage, there were 65,676 links between the SOPHID and HARS datasets (Table 3.14). This was after deduplicating 3,903 records from HARS where more than one record matched a single SOPHID record to leave 1,923 linked records. Similarly, 2,795 records from SOPHID, where more than one record matched a single HARS record, were deduplicated to leave 1,383 linked records. The algorithm linked 82.5% (65,676/79,596) of SOPHID records and 85.1% (65,676/77,212) of HARS records.

Almost all (97.6% [64,089]) of the records were linked with exact soundex code, date of birth and sex. Over half (55.1% [36,170]) of all links were corroborated by matching local patient ID number and clinic. A slightly higher percentage (61.2%) had matching risk group and ethnicity. Comparison of the records linked between XLatest and the SOPHID patient table showed that there were 3,804 (5.8%) for whom soundex codes did not match, 2,076 (3.2%) for whom dates of birth did not match, and 251 (0.4%) for whom sex did not match.

There were 367 (0.6%) links where the individual was reported to HARS as having died in a calendar year earlier than the individual was last reported to SOPHID. Over a third (36.8% [135]) of these 367 individuals were reported to HARS to have died in the year before they were last reported to SOPHID but over a quarter (26.7% [98]) were reported to have died more than five years before being last seen and 5.4% (20) more than ten years before.

For 26 of the 135 who had a date of death reported to HARS in 2006 but were last reported to SOPHID in 2007, the dates of death were thought to be incorrect. However, for the remaining 109 of the 135 records, the date of death appeared to be correct and was sometimes also reported to SOPHID, but a SOPHID report was inappropriately received for these individuals in the year after their death.

Examination of the latter 20 links with date of death in 1995 or 1996 indicated that all links between HARS and SOPHID were valid. Either the date of death was incorrectly reported (sometimes to both HARS and SOPHID), records were incorrectly deduplicated in SOPHID, the individual was subsequently reported by the clinic many years after their death (sometimes continuously), or the clinic had subsequently reported the same patient identifiers for a different individual.

Investigation of many of the remaining 212 records with more than one year difference between the year of death reported to HARS and the last year reported to SOPHID indicated that the date of death in HARS was incorrect (probably incorrectly linked to an ONS death report).

Table 3.14. Results of record linkage (✓ = exact match, X = discordant or missing information, empty cells not considered).

Match Score	Sdex[1]	Dob[2]	Sex	Clinic	Local patient ID number	Risk Group	Eth[3]	Date of HIV diagnosis[4]	Duplicates identified in HARS		Duplicates identified in SOPHID		Individual links after deduplication	
									n	Cumulative	n	Cumulative	n	Cumulative
1	✓	✓	✓	✓	✓				580	580	297	297	34,718	34,718
2	✓	✓	✓	one or other not matching					730	1,310	495	792	29,371	64,089
3	X	✓	✓	✓	✓				305	1,615	366	1,158	919	65,008
4	✓	X	✓	✓	✓				208	1,823	186	1,344	533	65,541
5	✓	✓	X					✓	31	1,854	12	1,356	64	65,605
6	✓	✓	X			✓	✓	Year of diagnosis only	69	1,923	27	1,383	71	65,676

[1] Soundex, [2] Date of birth, [3] Ethnicity, [4] Date of diagnosis in HARS, datepos in SOPHID

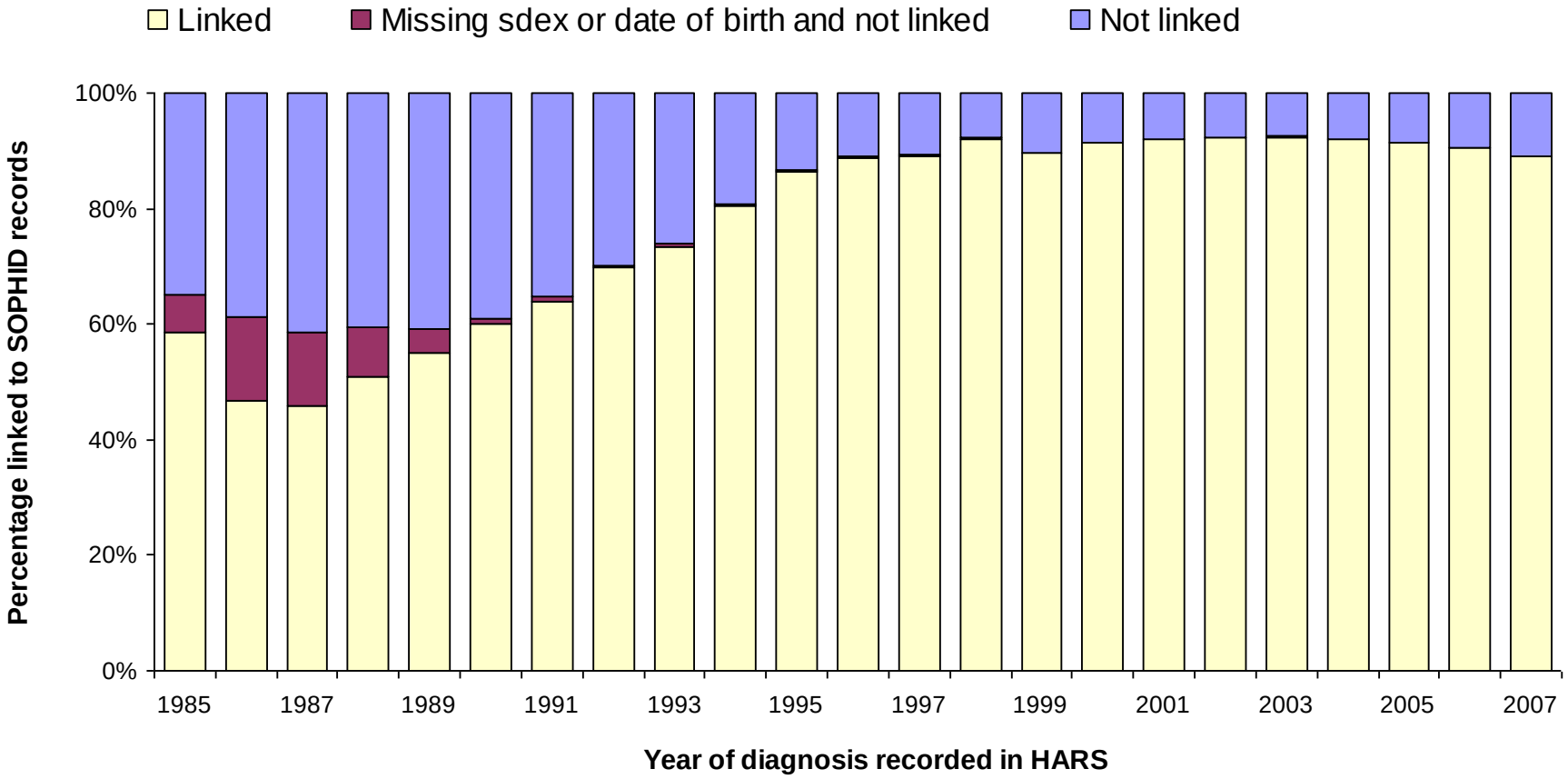
There were also 1,449 (2.2%) records where the individual was reported to SOPHID in an earlier year than they were reported as newly diagnosed to HARS. The largest percentage (45.5% [659]) of these records were only reported to SOPHID one calendar year before they were reported to HARS but the difference was greater than five years for 18.2% (264) and greater than ten years for 2.6% (38). Examination of a subset of these records indicated that these were reliable links. Where the difference was one year the date of diagnosis was reported late to HARS, possibly because the date of report was written on the form instead of the date of diagnosis. Where the difference was greater, the individual was likely to have attended different clinics over time and either the original diagnosis was not reported to HARS or the original diagnosis was reported but was not linked to the subsequent report to HARS. Of the 1,449 links between HARS and SOPHID records, there were 42 where the year of last report to SOPHID was before the year of HIV diagnoses reported to HARS. These also appeared to be reliable links with the same causes as described above. However, a third (14) of these HARS records were reports of deaths compared to only 63 (4.3%) of these 1,449 links.

The factors associated with record linkage were reassessed after this final process and compared to the previous assessment of record linkage between HARS and SOPHID carried out in 2003. As previously, the most significant factor was the period of HIV diagnosis, which was associated with a substantial increase in the proportion linked over time (Table 3.15). There was a similar pattern of record linkage to that in 2003 but with a higher proportion linked in 1995 and 1996, due to inclusion of SOPHID records from those years, and a

Table 3.15. Number and percentage of HARS records linked to SOPHID

	Number of linked HARS patients n (distribution, %)	All newly diagnosed patients n (distribution, %)	Percentage of newly diagnosed patients linked %	Multivariable adjusted odds ratio (95% confidence interval)	2003 multivariable adjusted odds ratio (95% confidence interval)
Period of HIV diagnosis					
1979-1984	322 (0.5)	431 (0.6)	75	0.17 (0.12, 0.23)	0.18 (0.14, 0.24)
1985-1989	3,163 (5)	6,146 (8)	51	0.15 (0.14, 0.17)	0.12 (0.11, 0.13)
1990-1994	6,807 (10)	9,684 (13)	70	0.30 (0.27, 0.33)	0.24 (0.22, 0.27)
1995-1999	11,361 (17)	12,731 (16)	89	0.62 (0.56, 0.68)	0.77 (0.71, 0.84)
2000-2004	25,325 (39)	27,509 (36)	92	0.83 (0.77, 0.90)	-
2005-2007	18,698 (28)	20,711 (27)	90	-	n/a
Age group					
15-24	7,994 (12)	9,663 (13)	83	1.16 (1.05, 1.27)	1.07 (0.98, 1.17)
25-29	13,762 (21)	16,407 (21)	84	1.07 (0.99, 1.16)	1.04 (0.96, 1.12)
30-34	15,494 (24)	18,052 (23)	86	-	-
35-39	12,214 (19)	14,112 (18)	87	1.00 (0.91, 1.08)	0.93 (0.85, 1.01)
40-44	7,305 (11)	8,492 (11)	86	0.93 (0.84, 1.03)	0.88 (0.80, 0.98)
45+	8,907 (14)	10,486 (14)	85	0.73 (0.67, 0.80)	0.82 (0.74, 0.90)
Risk group					
Sex between men	28,641 (44)	33,369 (43)	86	-	-
Heterosexual sex (men)	12,312 (19)	14,184 (18)	87	0.59 (0.53, 0.64)	0.51 (0.47, 0.55)
Heterosexual sex (women)	20,565 (31)	22,902 (30)	90	0.73 (0.67, 0.80)	0.58 (0.54, 0.63)
Injecting drug use	2,176 (3)	3,108 (4)	70	0.56 (0.49, 0.63)	0.46 (0.41, 0.51)
Blood/blood products	683 (1)	846 (1)	81	**1.19 (0.93, 1.52)**	**1.91 (1.53, 2.40)**
Other/not known	1,299 (2)	2,803 (4)	46	0.36 (0.31, 0.40)	0.17 (0.15, 0.20)
Region of diagnosis					
London	34,953 (53)	41,755 (54)	**84**	-	-
Outside London	30,723 (47)	35,457 (46)	**87**	**0.76 (0.72, 0.80)**	**1.16 (1.10, 1.23)**
Ethnicity					
White	32,707 (50)	34,585 (45)	95	-	-
Black African	24,550 (37)	26,690 (35)	92	0.53 (0.48, 0.58)	0.71 (0.64, 0.79)
Black Caribbean	2,158 (3)	2,318 (3)	93	0.60 (0.51, 0.72)	0.71 (0.57, 0.88)
Black Other	874 (1)	937 (1)	93	0.64 (0.49, 0.83)	0.62 (0.42, 0.91)
Indian\Pakistani\Bangladeshi	896 (1)	987 (1)	91	0.53 (0.42, 0.67)	0.57 (0.43, 0.76)
Other/mixed	2,791 (4)	2,948 (4)	95	0.77 (0.65, 0.92)	0.63 (0.50, 0.78)
Not known	1,700 (3)	8,747 (11)	19	0.02 (0.02, 0.02)	0.29 (0.27, 0.31)
Total	65,676 (100)	77,212 (100)	85	-	-

Figure 3.8. Percentage of HARS records linked to SOPHID records by year of diagnosis (1985 to 2007).



generally higher proportion linked in earlier years (Figure 3.8). Individuals infected through blood/blood products were no longer significantly more likely to have records linked than individuals infected through sex between men (p=0.17) as record linkage improved little for this group. Record linkage in 2008 was greater for individuals diagnosed in London than elsewhere, in contrast to 2003. The effects of age group and ethnicity were similar to 2003.

3.5.3.1 Discussion

This final record linkage process was relatively non-labour-intensive and quick because the algorithm was restricted to matches that were considered very robust. Yet the yield was improved and a number of additional duplicates were identified in each of the datasets. Incorporation of validations to the matching prevented many incongruous links being made between individuals where one was reported to have died before the other was reported. There were some inconsistencies between dates reported to HARS and SOPHID but examination of individual records indicated that the links were valid and reliable. Reports to SOPHID before dates of HIV diagnosis in HARS were valid and therefore the earliest date seen for care should be used as the date of HIV diagnosis of the individual in the integrated dataset. However, links between records that included a report of death to HARS before the last report to SOPHID should not be used in the integrated dataset because the data indicated that either the date of death was wrong or the individual was inappropriately reported to SOPHID after their death. These could be used in future analyses but would require manual checking and amendment to ensure data validity.

3.5.4 Summary of record linkage between HARS and SOPHID

Overall, approximately 75-80% of potential matches between HARS and SOPHID were highly likely to be true links (excluding fuzzy matches). With just soundex, date of birth and sex, the algorithm linked 73% of the 1997-2002 SOPHID records to HARS but 83% of the 1995 to 2007 SOPHID records to HARS. Fuzzy matching increased the percentage of the 1997-2002 records linked by 5% (to 78%) whereas matching with local patient ID number increased the percentage of the 1995-2007 records linked by 2% (to 85%). Overall, there were 28,680 records linked in data to the end of 2002, 36,143 to the end of 2003 and 65,676 to the end of 2008. Investigation of inconsistent dates resulted in the exclusion of fuzzy soundex and date of birth matching in the final algorithm and exclusion of links where the year of death was inconsistent with other reported dates. After these exclusions, a final total of 65,309 records linked between HARS and SOPHID were accepted for the integrated dataset.

This latest analysis supported the previous conclusion that the majority of individuals diagnosed before 1995 with HARS records that did not link to SOPHID records had probably died without their death being reported. This was consistent with the decreasing proportion linked before 1995 (from 86.4% in 1995 to 46.6% in 1986) but also with the higher proportion matched in the early years after inclusion of SOPHID data from 1995 and 1996 (the last two years before HAART substantially reduced mortality). However, the decreasing proportion linked was also likely to reflect some emigration and name changes.

Between 1995 and 2007, when SOPHID data have been collected, an annual mean of 9.1% of new HIV diagnoses were not linked to SOPHID. Previous analysis suggested that a quarter to a third of these reports could be fuzzy

matched but reliable linkage would require manual review. It is likely that a further substantial proportion of the non-linked reports between 1995 and 2007 were related to the same individual but reported with identifiers too different to allow specific record linkage without extensive manual review. It is therefore likely that only a minority of individuals who have positive test results never return for results or subsequent care.

There is a minimal cost associated with record linkage between HARS and SOPHID as the algorithms can be run quickly using existing data management resources. One-off exact matching for specific analyses would be the cheapest process as it requires only a simple, ad-hoc database query but it would only link 83% of historical records. Routine use of exact matching for record linkage at little extra cost could reduce costs of follow-up, improve data consistency and provide regularly updated linked datasets for analysis. The more complex record matching techniques developed and validated for this thesis can be run almost as quickly and simply as exact matching queries, provide some additional yield, and exclude probable false links to further improve data quality. Therefore, different record linkage methodologies are probably of comparable cost-effectiveness and the new algorithms developed may not have substantially reduced potential inclusion bias in subsequent analyses as yields are not greatly improved. However, the understanding of the data for future analyses should provide a long-term benefit for the improved use and interpretation of surveillance data. In particular, it is useful to know the limits of record linkage and that non-linked records are not necessarily due to under-reporting. Furthermore, cost-effectiveness of surveillance is likely to improve if routine processes are developed around a core system of record linkage.

3.6 Record linkage of SOPHID and CD4 records

3.6.1 Use of local patient ID numbers in the 1997-2003 SOPHID dataset to link to CD4 records (2004)

3.6.1.1 Data preparation

The 1997-2003 SOPHID dataset was used (Section 3.5.2.1), which contained the local patient ID number requested for all individuals reported to the 2003 survey. Records of patients reported in more than one year were linked on the basis of matching soundex code, date of birth and sex. Matching reports that were known to relate to more than one individual were not linked. There were 158,459 records for the 49,197 individuals in the SOPHID 1997-2003 dataset.

The CD4 Surveillance database archived in January 2004 contained 69,498 records of individuals with CD4 counts aged older than 14 years. This dataset contained 17,597 records with no soundex code, 6,152 records with no date of birth (2,892 records with neither soundex nor date of birth) and 6,838 records with no sex. There were 36,027 records that had been linked to records in HARS. A subset of 4,491 records had no local patient ID number reported from the laboratory, of which 402 were reported with only a soundex but no date of birth and 27 were reported with only a date of birth but no soundex.

3.6.1.2 Record linkage process

A different strategy was used than that used for linking reports between the SOPHID and HARS databases in 2004 (Section 3.5.2.2). Linked records were not excluded from further matching because it was thought that duplicates could be identified as multiple links to the same record. However, it was necessary to ensure that links were mutually exclusive so that the same pair of records was

not counted more than once. Therefore, a record of all links was made where more than one SOPHID record linked to a CD4 record or vice versa. There were up to three links per record and these were defined as primary, secondary and tertiary links according to the amount of information that matched. Primary links had more matching information than tertiary links indicating that the records were more likely to relate to the same individual. Arbitrary allocation was made where more than one SOPHID record linked to a CD4 record with an equivalent degree of matching information or vice versa. Data that were missing in both datasets were not considered as a match.

Fuzzy local patient ID numbers were where a five-digit/character element of the number was present in the matching local patient ID number. Because sites can use similar formats for their patient ID numbers, these were only used to link records if the dates of birth and location also matched. Fuzzy soundex and date of birth matches were not carried out because corroborating information on risk group and ethnicity was not available and the majority of fuzzy matches were expected to have been identified through local patient ID number and location matching. Because laboratories were the source of CD4 data whereas hospitals/clinics were the usual source of SOPHID data, an intermediary table was created to match comparable data sources. Laboratories, hospitals and clinics in similar geographical areas were grouped taking account of reported sources of CD4 counts and prior knowledge.

Visual FoxPro 5.0 was used for the analysis.

3.6.1.3 Results

There were 33,406 primary links made between the SOPHID and CD4 datasets (Table 3.16). These accounted for 68% (33,406/49,197) of the SOPHID records and 48% (33,406/69,498) of the CD4 Surveillance records. Of the primary records linked, 26,585 (80%) were exact soundex, date of birth and sex matches and a further 992 (3.0%) had the same soundex and date of birth but different sex. There were 17,312 (52% of the total) primary links made using local patient ID numbers of which 4,453 were fuzzy matches. Of the latter, 3,081 (69%) also matched on soundex and date of birth compared to 8,402 (65%) of the 12,859 exact local patient ID number matches. There were 1,234 and 653 links where the soundex and date of birth respectively were discordant and 3,968 and 32 where the CD4 soundex and date of birth respectively were missing. Record linkage to SOPHID was achieved for 3,971 (23%) of the 17,597 CD4 records with no soundex code but only 389 (5.7%) of the CD4 records with no sex and only 32 (0.5%) of the 6,152 CD4 records with no date of birth.

There were 959 secondary links and 15 tertiary links indicating that 944 CD4 records linked to two SOPHID records and 15 CD4 records linked to three SOPHID records. In comparison, there were 1,080 SOPHID records linked to two CD4 records and 14 SOPHID records linked to three CD4 records (primary matching). This suggests that 2% (974/49,197) of SOPHID records and 1.6% (1,108/69,498) of CD4 records were potentially duplicates or that records from more than one individual were inappropriately deduplicated in each dataset.

Table 3.16. Results of record linkage between SOPHID and CD4 Surveillance (✓ = exact match, X = discordant)

Match Score	Soundex code	Date of birth	Sex	Loc[1]	Local patient ID number	Primary links		Secondary links	Tertiary links
						number	cumulative (%)		
1	✓	✓	✓	✓	✓	8,187	8,187 (25)	147	3
2	missing	✓	✓	✓	✓	2,653	10,840 (32)	31	
3	X	✓	✓	✓	✓	904	11,744 (35)	304	8
4	✓	X	✓	✓	✓	498	12,242 (37)	110	2
5	X	X	✓	✓	✓ >4 characters	243	12,485 (37)	29	
6	✓	✓	X or missing	X	✓	215	12,700 (38)	5	
7	X or missing	✓	X or missing	X	✓	159	12,859 (38)	20	
8	✓	✓	✓	✓	✓ >4 characters	3,026	15,885 (47)	24	
9	✓	✓	X	✓	✓ >4 characters	55	15,940 (48)	4	
10	X or missing	✓	X or missing	✓	✓ >4 characters	1,372	17,312 (52)	131	2
11	✓	✓	✓	✓	X	10,438	27,750 (83)	93	
12	✓	✓	X	✓	X	485	28,235 (84)	18	
13	✓	✓	✓	X	X	4,934	33,169 (99)	43	
14	✓	✓	X	X	X	237	33,406 (100)	0	

[1] Location

There were 33,471 (48%) records in CD4 Surveillance that had not previously been linked to records in HARS. Record linkage to SOPHID confirmed that 17% (5,791) of these records were derived from HIV-infected individuals. Many of these records were probably not previously linked to HARS because they were missing soundex codes (43% [2,472]). Overall, record linkage of CD4 Surveillance records to SOPHID increased the proportion of records linked to either HARS or SOPHID from 52% to 60%. Very few records without a date of birth were linked to either HARS or SOPHID (<1%) although these records only accounted for 9% (6,152/69,498) of all CD4 Surveillance records.

3.6.1.4 Discussion

A substantial number and proportion of CD4 Surveillance records were linked to SOPHID records. Soundex, date of birth and local patient ID number were all essential to maximise the number linked. However, a substantial number of records with missing soundex were linked indicating that many of these would now link to HARS.

A considerable number of the linked records were not previously linked to HARS records therefore confirming that these individuals reported to CD4 Surveillance were HIV-infected. However, 40% of CD4 records remained unlinked. It was likely that many of these were duplicate records or records that could not be linked because they lacked soundex and/or date of birth information. The substantial proportion of records with both soundex and date of birth that remained unlinked were possibly from individuals who were no longer in E,W&NI in the years of the SOPHID surveys due to either death or emigration. It is also likely that a number of these records were from individuals without HIV infection (Section 2.3.5).

3.6.2 Record linkage between CD4 Surveillance and SOPHID using multiple records for each individual: 2008

As for record linkage between HARS and SOPHID, the datasets for record linkage between SOPHID and CD4 Surveillance were modified in 2008 although the central methodology remained the same. The datasets included all reports to CD4 Surveillance and SOPHID (Report Tables), which permitted record linkage between multiple reports of an individual with all different combinations of the patient identifiers considered for matching.

This process was carried out by a junior member of staff under my supervision.

3.6.2.1 Data sources

There were 1,306,513 records relating to 112,215 adults in the CD4 Surveillance dataset to the end of 2007 and 562,130 records relating to 79,596 adults (aged older than 14 years) in the SOPHID 1995-2007 dataset. Records were not excluded because of missing date of birth, soundex or sex as these records could still be linked despite missing information.

3.6.2.2 Record linkage process

Reports in the SOPHID and CD4 Surveillance databases were linked using a streamlined version of the previous record linkage algorithm (Table 3.17). At least two of soundex code, date of birth and local patient ID number were required to match. Matching dates of birth were accepted if at least five characters of the local patient ID number also matched (match score 5). Matching patient ID numbers from any sites were accepted if at least two of the day, month or year of birth also matched (match score 6). The location of the

clinic or laboratory was not considered as this had not been found to substantially improve the sensitivity or specificity of the algorithm. This algorithm also did not incorporate fuzzy soundex code and date of birth matching. Microsoft Access 2003 was used for the analysis.

As previously, the record linkage algorithm was not mutually exclusive and records were not excluded from subsequent matching. Therefore, records could be linked a number of times between the datasets, identifying possible duplicates. Where possible duplicate records were found, the link with the lowest match score was retained or arbitrary allocation was made if there was an equivalent degree of matching information.

3.6.2.3 Results

At the end of the record linkage process, there were 59,814 links between the SOPHID and CD4 Surveillance datasets (Table 3.17). This was after deduplicating 9,888 records from the CD4 Surveillance dataset and 739 records from SOPHID. The algorithm linked 75% (59,813/79,596) of all SOPHID records and 63% (59,813/94,873) of all CD4 Surveillance records (denominators updated to account for deduplication). Of the records linked, 50,819 (85%) had the same soundex code and date of birth, 48,275 (81%) had the same local patient ID number and 41,284 (69%) had all three matching. There were 8,994 (15%) links that had either a different, or missing, soundex code or date of birth. The majority of these (93% [8,346/8,994]) had discordant soundex codes. Only 2,002 (3.3%) of the links were based on fuzzy local patient ID number (match score 5) or fuzzy date of birth (match score 6).

Table 3.17. Results of record linkage between SOPHID and CD4 Surveillance datasets (✓ = exact match, empty cells not considered)

Match Score	Soundex code	Date of birth	Sex	Local patient ID number	Primary links	
					number	cumulative (%)
1	✓	✓		✓	41,284	41,284 (69)
2	discordant or missing	✓		✓	6,648	47,932 (80)
3	✓	✓	✓	discordant or missing	9,535	57,467 (96)
4	✓	one or other not matching		✓	344	57,811 (97)
5		✓		✓ >4 characters	1,699	59,510 (99)
6		✓ two of day, month or year		✓	303	59,813 (100)

3.6.2.4 Factors associated with record linkage

The factors associated with record linkage between CD4 Surveillance and SOPHID were similar to those between HARS and SOPHID. An increasing proportion of SOPHID records were linked to CD4 Surveillance with increasing year that an individual was first reported to SOPHID (Table 3.18). However, a lower proportion of records were linked in 2006 and 2007 than between 2000 and 2005. This may have been because a few laboratories were not able to report CD4 cell counts for 2006 and 2007 (Section 2.3.2). Individuals infected through sex between men were significantly more likely to have records linked than all other individuals and particularly more than for individuals without a reported risk group. Fewer records of black Africans or black Caribbeans linked than for other individuals except that markedly fewer records linked for individuals without a reported ethnicity. There was a substantially higher proportion of records linked for individuals who were first reported from London compared to elsewhere.

Table 3.18. Number and percentage of SOPHID records linked to CD4

	Number of linked SOPHID patients n (distribution, %)	All SOPHID patients n (distribution, %)	Percentage of SOPHID patients linked %	Multivariable adjusted odds ratio (95% confidence interval)
First year reported to SOPHID				
1995	10,070 (16)	13,467 (17)	75	**0.89 (0.84, 0.95)**
1996-1997	5,330 (9)	7,736 (10)	69	**0.69 (0.65, 0.74)**
1998-1999	5,256 (9)	7,112 (9)	74	**0.87 (0.81, 0.93)**
2000-2001	6,958 (12)	8,911 (11)	78	**1.22 (1.14, 1.30)**
2002-2003	10,472 (18)	13,598 (17)	77	**1.34 (1.27, 1.42)**
2004-2005	11,701 (20)	14,464 (18)	81	**1.72 (1.62, 1.82)**
2006-2007	10,026 (17)	14,308 (18)	70	-
Age group				
15-25	4,871 (8)	6,598 (8)	74	1.06 (0.99, 1.14)
25-29	10,576 (18)	13,960 (18)	76	1.04 (0.98, 1.10)
30-34	14,379 (24)	19,149 (24)	75	-
35-39	12,672 (21)	16,560 (21)	77	**1.07 (1.02, 1.13)**
40-44	7,689 (13)	10,227 (13)	75	1.01 (0.95, 1.07)
45+	9,626 (16)	13,102 (16)	73	0.97 (0.92, 1.03)
Risk group				
MSM	26,280 (44)	32,643 (41)	81	-
Heterosexual men	11,023 (18)	14,909 (19)	74	**0.75 (0.71, 0.79)**
Heterosexual women	18,222 (30)	24,185 (30)	75	**0.80 (0.75, 0.85)**
IDU	1,845 (3)	2,624 (3)	70	**0.66 (0.61, 0.73)**
Blood/blood products	713 (1)	1,129 (1)	63	**0.60 (0.52, 0.68)**
Other/not known	1,730 (3)	4,106 (5)	42	**0.18 (0.17, 0.20)**
Region first reported from				
London	35,097 (59)	42,424 (53)	83	-
Outside London	24,716 (41)	37,172 (47)	66	**0.33 (0.32, 0.34)**
Ethnicity				
White	30,715 (51)	39,615 (50)	78	-
Black African	21,068 (35)	28,491 (36)	74	**0.78 (0.74, 0.82)**
Black Caribbean	1,885 (3)	2,449 (3)	77	**0.78 (0.71, 0.87)**
Black Other	1,108 (2)	1,366 (2)	81	0.98 (0.85, 1.13)
Indian\Pakistani\Bangladeshi	785 (1)	1,009 (1)	78	1.09 (0.93, 1.28)
Other/mixed	3,028 (5)	3,772 (5)	80	0.97 (0.89, 1.06)
Not known	1,224 (2)	2,894 (4)	42	**0.28 (0.26, 0.30)**
Total	59,813 (100)	79,596 (100)	75	-

3.6.2.5 Discussion

This record linkage process was more efficient than the previous one because it used strategies that I had developed over time. In particular, the algorithm was restricted to robust matches based on at least two of soundex code, date of birth and local patient ID number and all reported combinations of those patient identifiers were considered for record linkage. This latest algorithm was not based on soundex with fuzzy date of birth or local patient ID number due to redundancy in soundex codes (Section 3.4.1). Yet, it produced a yield substantially higher than that achieved with a more complicated algorithm in 2004 (75% of SOPHID records linked versus 68% in the 2004 analysis). This may have been due in part, to improved reporting of local patient ID number to SOPHID from clinics outside London. A large number of duplicate CD4 Surveillance records were found, which provides further evidence of under-deduplication in that database due to incompleteness of patient identifiers (Chapter 2.3).

3.7 Conclusion

There was no gold standard to use for comparison of links between datasets. Although the majority of links had corroborating information and were thought to be true links it appeared that there was a substantial, but unknown, number of additional true links that could not be identified due to coding or other errors. The subset investigation of the non-exact matches showed that a small but significant proportion of true links were missed if only exact soundex, date of birth and local patient ID number were used. Fuzzy matching probably identified some of these true links but introduced errors by falsely linking individuals. Also, substantial numbers remained unlinked even after fuzzy matching.

Determination of which fuzzy matches were true links and identification of additional true links would require the allocation of considerable resources for follow-up. Therefore, links made with fuzzy matching were not used for record linkage in 2008 to create the final integrated dataset for subsequent analyses.

There were some significant differences between record linkage of HARS and SOPHID compared to SOPHID and CD4 Surveillance. Whereas the percentage of HARS records linked to SOPHID (85%) was similar to the percentage of SOPHID records linked to HARS (83%), the percentage of SOPHID records linked to CD4 Surveillance (76%) was substantially greater than the percentage of CD4 Surveillance records linked to SOPHID (63%). This was because there were a similar number of records of individuals in the SOPHID and HARS datasets used for record linkage but many more records of individuals in the CD4 Surveillance dataset. This is likely to be due to two main reasons: i) a reduced ability to deduplicate the CD4 Surveillance dataset because of incomplete reporting of soundex codes and, to a lesser extent, dates of birth ii) the reporting of CD4 cell count data for individuals who did not have HIV infection from laboratories unable to distinguish patients by HIV status. Because of missing soundex codes and dates of birth in the CD4 Surveillance dataset, the final SOPHID-CD4 Surveillance record linkage was based relatively less on these patient identifiers (85%) and more on local patient ID numbers (81%) than SOPHID-HARS record linkage (98% and 55% respectively). This also resulted in a higher percentage of different, or missing, soundex codes, dates of birth or sex between SOPHID and CD4 Surveillance links (15% [8,995/59,814]) than between SOPHID and HARS links (2.4% [1,587/65,676]).

Record linkage in 2008 resulted in 65,676 links between HARS and SOPHID and 59,814 links between SOPHID and CD4 Surveillance. Record linkage between HARS and SOPHID shows that it can enable ethnicity information to be shared between these complementary national HIV surveillance systems. This would considerably reduce the need for follow-up and the amount of missing information used in further analyses of data, strengthening the conclusions made. Also, the numbers of records linked were large enough to allow analysis of combined data such as the date of diagnosis from HARS, treatment information from SOPHID and the CD4 cell count at HIV diagnosis from CD4 Surveillance. However, analyses that make use of integrated datasets need to take account of bias introduced by record linkage.

Between both HARS and SOPHID and between SOPHID and CD4 Surveillance, the proportion of linked records was substantially affected by the year of HIV diagnosis, the risk group and ethnicity as well as whether the individual was first reported from London or not. Significantly fewer records were linked between HARS and SOPHID for individuals diagnosed in the years before 1995, indicating that many of these individuals were likely to have died without being reported, emigrated or changed names (such that they did not appear in SOPHID records in 1995). The under-reporting of deaths was probably substantial prior to 1995 as the death rate was high and ONS reports of non-HIV-related deaths were not received until 1997. This is supported by the reduced record linkage for older individuals who were more likely to have died. The poor record linkage for individuals diagnosed during these early years disproportionally affected MSM, IDU and individuals infected through blood/blood products as a much higher proportion of heterosexuals were

diagnosed since the mid-1990s. Of those who were not reported to have died before 1995, 89.0% of heterosexuals were diagnosed after 1994 compared to 58.3% of MSM, 48.8% of IDU and 19.5% of individuals infected through blood/blood products. Due to the commonality between risk group and ethnicity, similar differences were observed by ethnicity though these effects were reduced after accounting for changes over time. This may be due to the unfamiliarity of uncommon names. Additionally, the proportion linked differed between reporting sites but little of this variation was captured by either their size or location (London or not). It is likely that the completeness and quality of reporting was more a function of the staff and their local patient record systems.

Record linkage between SOPHID and CD4 Surveillance was also affected by the same factors mentioned above. However, the proportion of records linked declined in the most recent years. This may be because not all of these recently diagnosed individuals had a CD4 cell count measurement by the end of 2007 or because some of their CD4 cell counts measurements had not yet been reported. As with record linkage between HARS and SOPHID, fewer records linked for heterosexuals and individuals of black African or black Caribbean ethnicity after adjusting for the year of first report to SOPHID (non significant difference for individuals of other non-white ethnicities). Record linkage was significantly greater for reports from London compared to those from outside London, which is likely to be due to the greater dependence on local patient ID number for matching and the longer inclusion of local patient ID number in the SOPHID data from London.

Record linkage has been used between reports of HIV/AIDS cases and deaths to quantify under-reporting or for identification of losses to follow-up. Probabilistic record linkage between known HIV/AIDS deaths and known non-deaths with active follow-up was assessed in Australia and found to have a sensitivity* of 82% and a specificity of 92%[229]. The limited sensitivity was attributed to a lack of patient identifiers on the HIV/AIDS databases. Various deterministic record linkage strategies between an AIDS registry and a hospital discharge file in New York State achieved sensitivities of 32-85% and positive predictive values of 14%-99%[230]. Errors in patient identifiers were also found to be a key limiting factor in the effectiveness of the process. These reporting errors are important to consider if record linkage is used to estimate under-reporting of HIV/AIDS cases because non-linked records cannot be considered to separate individuals if identifiers are not complete or reliable. This may have resulted in under-estimation of reporting completeness in Italy, France and E,W&NI[155;172;213;231]. Adjustments for under-reporting are no longer used in E,W&NI because results from this thesis have shown the frequent variability in reported patient identifiers for the same individual and the inability of deduplication within datasets and of record linkage between datasets to identify all records from the same individual.

The development of the record linkage algorithms took a great deal of time over a number of years. At the most complex, this involved a 72-part algorithm to

* Sensitivity measures the percentage of true positives that are correctly identified whereas specificity measures the percentage of true negatives that are correctly identified. For record linkage, sensitivity describes the percentage of records that should link between two datasets that are actually linked by a record linkage process. In contrast, specificity describes the percentage of records that should not link between two datasets that remain unlinked after the record linkage process. In this case, the true status of the patients had been confirmed through active follow-up against which record linkage strategies could be compared.

delineate all of the exact and fuzzy matches between HARS and SOPHID datasets, each of which required filters to be created on each of the datasets and joins made between them. Yet, amalgamation of accepted matches, which had previously been evaluated, resulted in much quicker processes. Exclusion of fuzzy matches, which were shown to increase the yield but with a need for manual record inspection, also increased the efficiency. The matching of multiple records per individual more than compensated for the exclusion of fuzzy matches without loss of specificity. Lessons learnt from previous record linkage resulted in the most efficient and productive process being implemented in 2008. However, ongoing manual checking of non-matching individual records and follow-up is recommended for continuing the integration of the surveillance systems and for future development of the record linkage algorithm.

Record linkage of HIV case reports is increasingly being undertaken by the HIV surveillance team to reduce the follow-up needed for missing information, to account for under-reporting and to integrate data for new analyses. Reducing follow-up lightens the workload of data providers but redirects the workload of the HIV surveillance team into record linkage instead of follow-up. However, record linkage needs to be reliable if used to update patient records without increasing the misclassification bias of subsequent analyses. I concluded that at least two of soundex, date of birth and local patient ID number should match for record linkage to be accepted without further manual review.

The supplementation of data from one surveillance system to another can result in increased accuracy through reductions in under-reporting or over-counting (due to duplicate records). If either of these effects is significant, published trends over time could be markedly affected. The inability to deduplicate all CD4

Surveillance records due to lack of patient identifiers and the potential reporting of non-HIV-infected patients means that non-linked reports should not be considered as under-reporting to HARS or SOPHID. Yet, record linkage raises the potential of supplementing HARS with SOPHID records and vice versa.

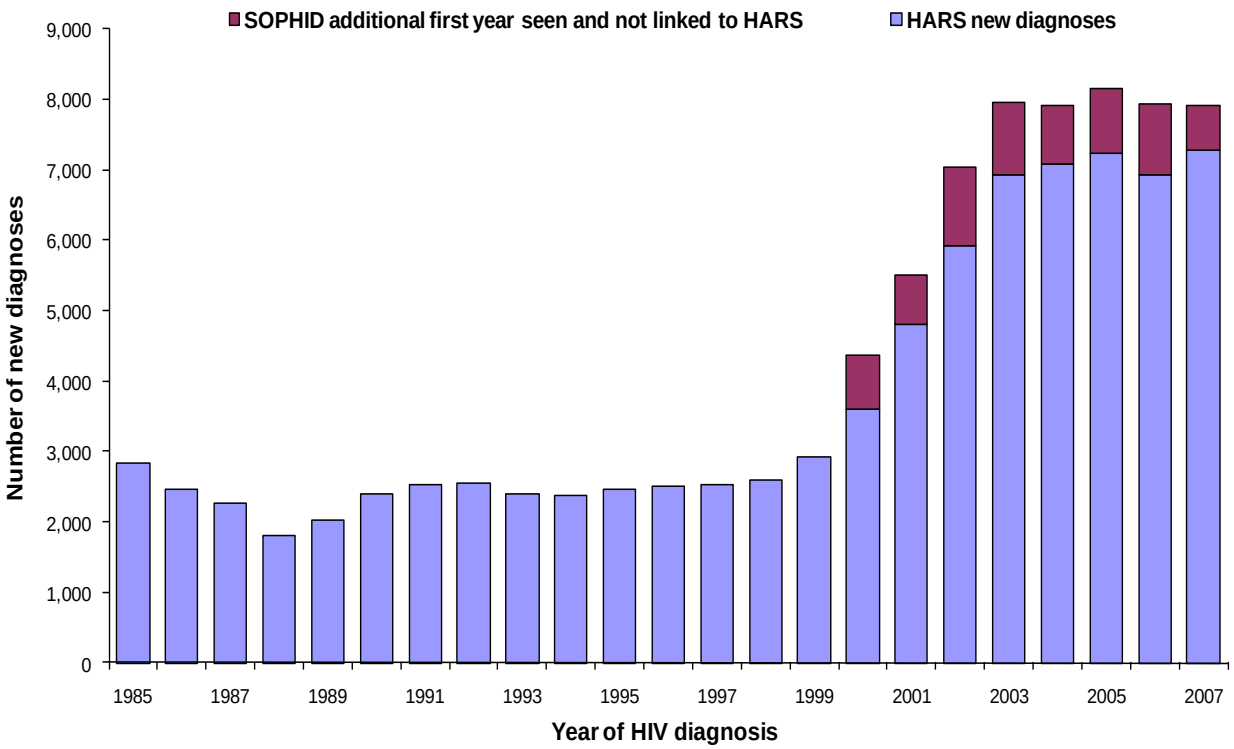
Based on results in this chapter, HPA surveillance systems are changing to incorporate ongoing record linkage and validation of HIV case reports as they are received. Since 2008, providers of SOPHID data were being prompted to clarify whether reports of HIV diagnoses that had not been received by SOPHID (according to record linkage using soundex, date of birth and local patient ID number) should in fact be added to SOPHID. Of the 83,393 HIV diagnoses reported in 2007, 853 were followed-up in this way but only 56 were added to SOPHID (276 were already reported with different identifiers and 510 had never returned for their results [the remainder should not have been reported according to the SOPHID protocol]). This indicated that there was very little under-reporting to SOPHID of individuals who were reported to HARS. This process will be continued because it is largely automated within the general SOPHID validation process and therefore requires minimal additional workload. Similar follow-up could be incorporated into CD4 Surveillance processes to improve record linkage and into HARS processes to improve completeness.

There were 6,993 SOPHID records for individuals first seen for HIV care between 2000 and 2007 that were not linked to HARS (Figure 3.9). Addition of these to HARS would increase the number of HIV diagnoses each year by a mean of 12.6% (874) during this period. The usual data entry process for addition of these records to HARS would be likely to find most of the non-linked fuzzy matches. The usual deduplication process would reduce the number of

records considered to be additional HIV diagnoses but it is likely that the result would still indicate significant under-reporting to HARS.

Substantial under-reporting to HARS, however, is not indicated by the very similar trends in the number of HIV diagnoses reported to HARS and the number of individuals first seen for clinical care in SOPHID (Figure 3.10). Numbers were slightly higher for SOPHID reports than HARS, which may indicate better deduplication in HARS, deduplication to reports of HIV diagnoses prior to 1995 in HARS, or under-reporting to HARS. Yet, the records that were linked accounted for 88.4% of the HARS records and 86.0% of the SOPHID records. The very similar patterns may provide further evidence that the non-linked records related to the same individuals and that these were reported to both systems but with different identifiers. It is unlikely that the effects of under-reporting and deduplication in each system would match each other so closely.

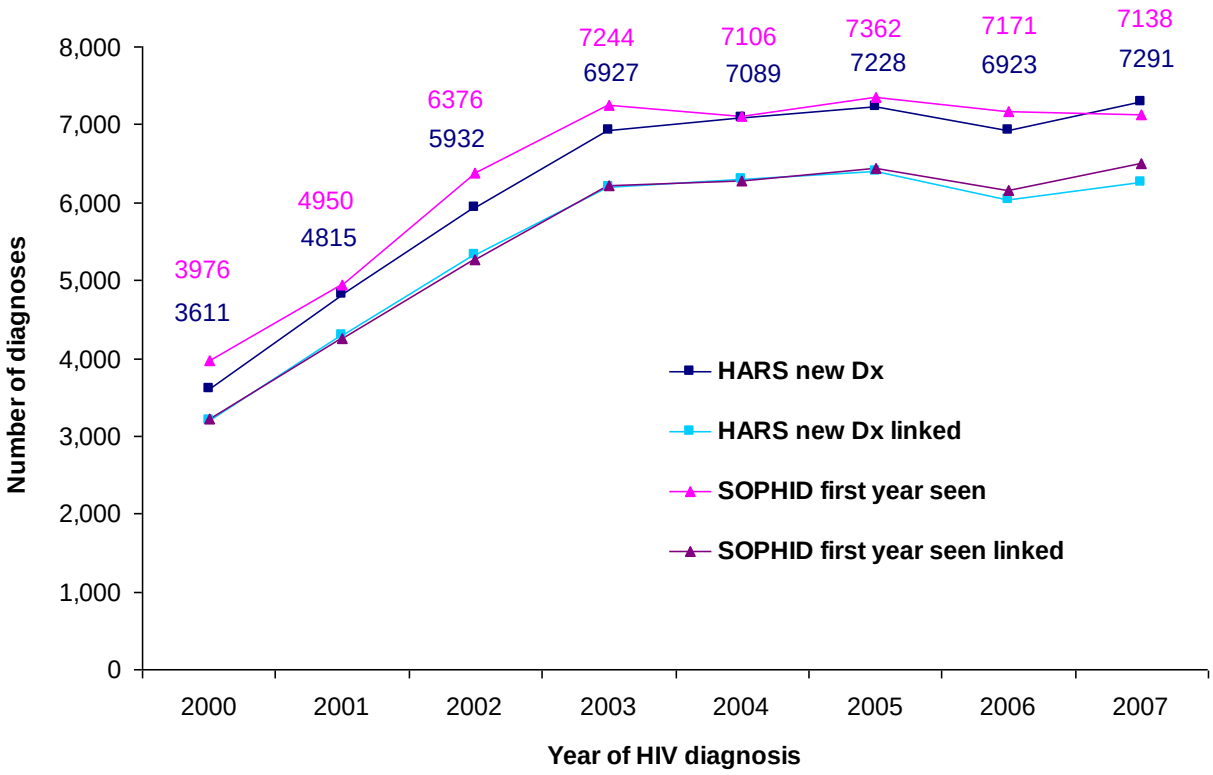
Figure 3.9. Numbers of HIV diagnoses over time if SOPHID records for individuals first seen for care that did not link to HARS were considered to be under-reported and added to the number of HIV diagnoses reported to HARS



In conclusion, to prevent over-counting due to the duplication of reports with different identifiers, the addition of non-linked SOPHID records to HARS should not be undertaken without further follow-up with the data providers. It is likely that many of these records will have also been reported to HARS but with patient identifiers too different to allow reliable record linkage. Although some under-reporting would be accounted for, this would inflate the numbers newly diagnosed. Multiple reporting systems have the potential to increase the ascertainment of reports but a single integrated surveillance system with unified resources for follow-up is likely to be more efficient and contain fewer inconsistencies, particularly if these systems are known to be subject to coding

errors. Therefore, there is a need for care when adding reports from novel, parallel surveillance systems (such as surveillance of recent HIV infections) to HARS because of the potential for creation of additional duplicates.

Figure 3.10. Numbers of HIV diagnoses reported to HARS and numbers first seen for clinical care reported to SOPHID



Finally, there is potential to link HIV surveillance data to patient records in other national HIV databases (of ART resistance, seroconverters and HIV-infected pregnant women), in research databases (such as the Collaborative HIV Cohort[232]), in other infectious disease surveillance databases (for example: sexually transmitted infections [STIs], TB, hepatitis, pneumococcal infection), and in non-infectious disease databases (such as cancer registries). Some of these have already been undertaken (resistance, TB, pneumococcal, hepatitis), are being developed (HIV-infected pregnant women [data from the National

Study of HIV in Pregnancy and Childhood]) or are being considered (Collaborative HIV Cohort and HPA STI databases). The methodologies have all been deterministic but the hierarchical algorithms have varied according to the commonality of data variables. Record linkage allows new analysis without additional and reproduced data collection and is therefore likely to be increasingly used in both surveillance and research as computing power increases, data transfer becomes more user-friendly and data management skills become more prevalent. In particular, data collection from NHS providers is becoming increasingly rationalised by central bodies (NHS Information Centre) to reduce the repeated reporting of the similar information. This is creating a need for further record linkage to integrate reports to achieve the vision of a single, centrally-mandated electronic care record for patients (National Programme for Information Technology). Both locally in HIV surveillance, and externally, statistics generated from these processes will need to put measures in place to prevent over-counting through duplication and to account for inclusion bias resulting from differential record linkage.

Chapter 4. Creation of a single, robust integrated dataset

4.1 Introduction

The primary aim of the processes described in this chapter was to generate and describe a single, robust integrated dataset from the records in each of the three independent HIV surveillance systems based on the bilateral record linkage achieved in Chapter 3. This dataset would then be used for all subsequent analyses. Creation of the integrated dataset required a decision of whether to exclude records that did not link uniquely between all three systems and selection and rationalisation of a common set of data variables to create patient records containing a coherent sequence of events.

The first part of this chapter focuses on the justification for the inclusion of records in the integrated dataset. The primary methodology was triangulation of records linked between SOPHID, HARS and CD4 Surveillance – my aim was to create unique and consistent links between all three datasets. This required consideration of links that did not form a coherent triangle, links that only formed two sides of the triangle, and multiple links to the same record.

Full integration of the three independent, national HIV case reporting systems depended not only on record linkage of the datasets but on integration of data variables. Information relating to the same HIV-infected individual from the different systems needed to be assimilated and validated so that a complete and consistent description of the individual with a coherent timeline of events was available for analysis. Data were examined to determine which should be used in the integrated dataset according to whether any inconsistencies were thought to be due to coding errors, under-reporting or unreliable record linkage.

Information selected for the integrated patient record, such as the date of diagnosis, first CD4 cell count, date first seen for clinical care, date of starting ART, date of AIDS diagnosis and date of death were then validated for chronological consistency. This process led to further exclusion of a small number of records but resulted in a final, single, robust, integrated dataset.

4.2 Aims

a) To create unique and consistent links between the three independent HIV surveillance datasets to generate a single, robust integrated dataset.

b) To integrate data variables from the three independent HIV surveillance datasets so that a complete and consistent description of the individual with a coherent timeline of events was available for analysis

4.3 Triangulation of record linkage between HARS, SOPHID and CD4 Surveillance to create the integrated dataset

4.3.1 Records linked between SOPHID, HARS and CD4 Surveillance

Triangulation of record linkage between HARS, SOPHID and CD4 Surveillance was carried out to create an integrated dataset. Links between SOPHID and HARS were considered as the basis for triangulation because the greater completeness of patient identifiers in these datasets conferred greater reliability to these links than those to CD4 Surveillance (Chapter 3). My ultimate aim was to obtain unique and consistent links between all three datasets (Figure 4.1a). However, in order to achieve this, it was necessary to find an approach that

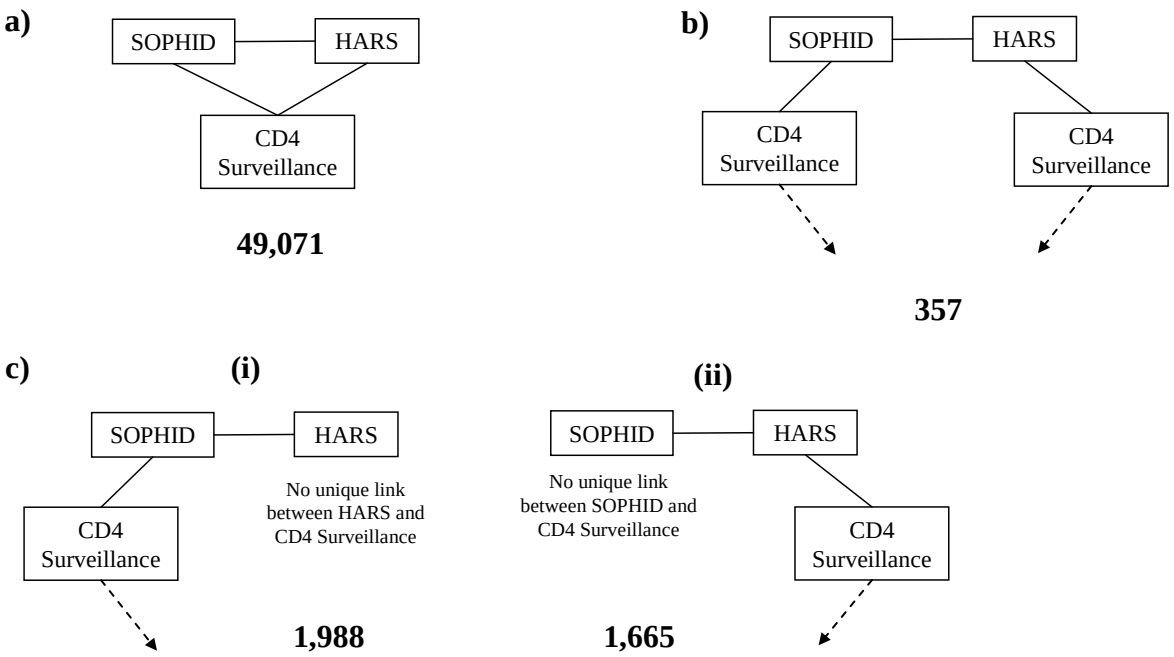
could be used where the three links did not triangulate (Figure 4.1b), or where there were only two links (Figure 4.1c). Where the three links did not triangulate, the CD4 Surveillance records were considered duplicates and these records were merged. Where there were only two links, the records were compared to see whether the non-linked records were likely to have been reported from the same individual, in which case the triangle could be completed. However, it was possible that some of the CD4 Surveillance records would link onwards to SOPHID or HARS records that were different to those forming the base of the triangle (Figures 4.1b and 4.1c). For example, where there were no links between HARS and CD4 Surveillance but links between SOPHID and HARS and between SOPHID and CD4 Surveillance, the records from CD4 Surveillance could link onwards to SOPHID, HARS or both SOPHID and HARS records (Figure 4.1d). Multiple SOPHID or HARS records linking to CD4 Surveillance could identify further duplicates or be unreliable, due to unspecific deduplication of CD4 Surveillance records.

There were 65,309 unique links between HARS and SOPHID after excluding 367 records where the date of death reported to HARS was before the date last seen for care as reported to SOPHID (Section 3.5.2.4). Of these, there were 49,071 unique links between HARS, SOPHID and CD4 Surveillance (Figure 4.1a). There were 357 links between HARS and SOPHID that each linked to a CD4 Surveillance record but where these CD4 Surveillance records were different (Figure 4.1b). There were 53,605 links where SOPHID records linked both to HARS and CD4 Surveillance but for 1,988 of these, the HARS record did not uniquely link to CD4 Surveillance (Figure 4.1c(i)). Conversely, there were 49,696 links where HARS records linked both to SOPHID and CD4

Surveillance but for 1,665 of these, the SOPHID record did not uniquely link to CD4 Surveillance (Figure 4.1c(ii)).

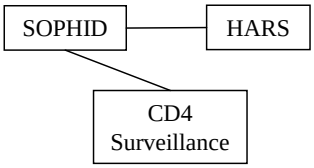
The 357 links shown in Figure 4.1b were considered to be due to duplicates in the CD4 Surveillance database because they linked separately to HARS and SOPHID. These were records of individuals that had information reported differently at the time of HIV diagnosis and during follow-up (most had been diagnosed prior to data collection by SOPHID) and with some information missing from reports to CD4 Surveillance (e.g. Figure 4.2)

Figure 4.1. Schema showing combinations of linked records between HARS, SOPHID and CD4 Surveillance (number of records shown in bold)



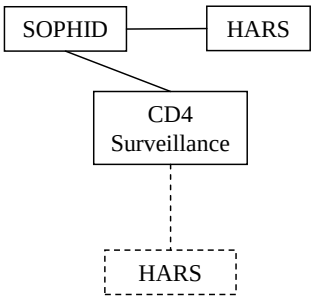
d)

(i)



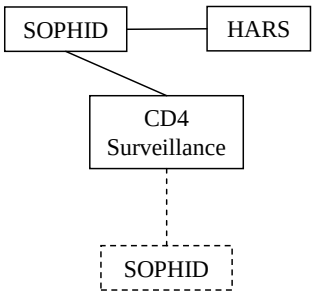
1,594

(ii)



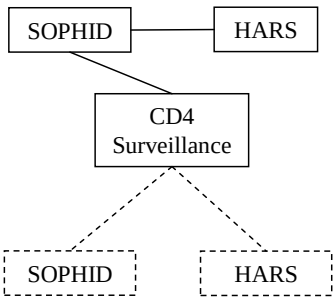
173

(iii)



25

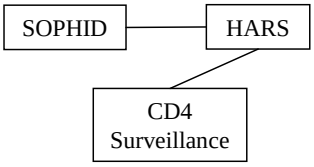
(iv)



196

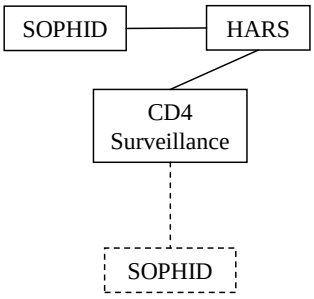
e)

(i)



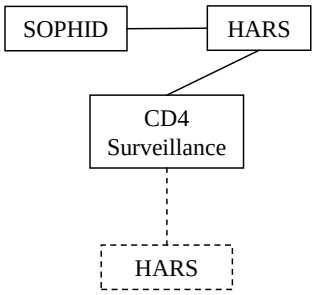
947

(ii)



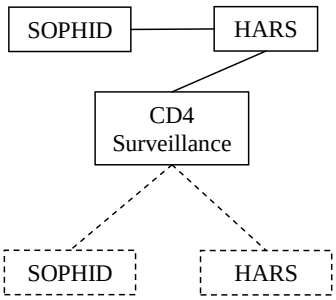
158

(iii)



28

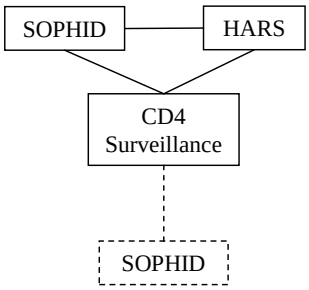
(iv)



532

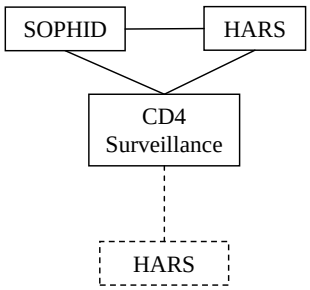
f)

(i)



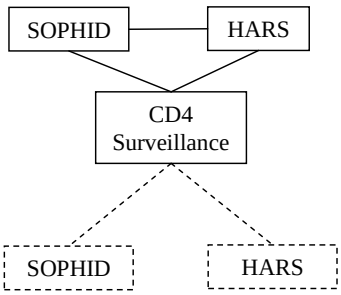
169

(ii)



2,086

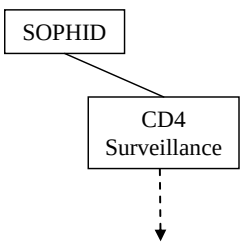
(iii)



711

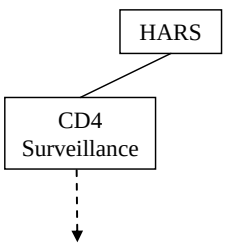
g)

(i)



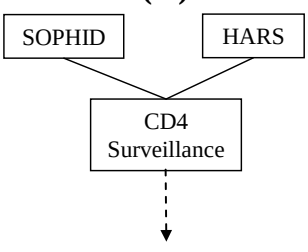
6,467

(ii)



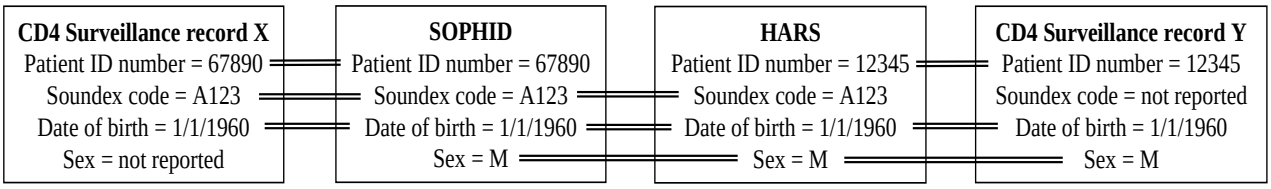
4,723

(iii)



734

Figure 4.2. Deduplication of CD4 Surveillance records through triangulation



The links shown in Figure 4.1c(i) and 4.1c(ii) were made up of combinations of links shown in Figure 4.1d and 4.1e respectively. These links were considered in the order described above to see whether the records could be reliably triangulated, either through further deduplication of records (Figures 4.1b, 1d(ii), 1d(iii) 1e(ii) and 1e(iii)), or through linkage of the third side of the triangle (Figures 4.1d(i), 1d(iii), 1e(i) and 1e(iii)). The 196 and 532 records with multiple links to both HARS and SOPHID (Figure 4.1d(iv) and Figure 4.1e(iv)) were excluded from further analysis as these were determined to include a high proportion of records that had been inappropriately deduplicated in the CD4 Surveillance dataset. The 1,594 and 947 links in Figure 4.1d(i) and Figure 4.1e(i) were found to be fuzzy matches with either missing or discordant information between HARS/SOPHID and CD4 Surveillance and it was therefore thought that the third side of the triangle could be linked with confidence. The 173 and 158 links in Figure 4.1d(ii) and 4.1e(ii) were found to consist of a mixture of duplicates and non-duplicates and were therefore excluded from further analysis. The 25 and 28 links in Figure 4.1d(iii) and 4.1e(iii) were excluded as there was no evidence that the multiple SOPHID or HARS records were duplicates.

The triangulation process above resulted in a total of 51,969 unique links between all three datasets (Figures 4.1a + 1b + 1d(i) + 1e(i)). There were a total of 4,078 further records that linked but that were excluded from further analysis as they included multiple links between datasets, largely due to records that had been inappropriately deduplicated in the CD4 Surveillance dataset. There remained 9,262 records linked uniquely between HARS and SOPHID where neither was linked to CD4 Surveillance records.

Finally, there were 6,467 SOPHID records that were linked to CD4 Surveillance records but not to HARS records (Figure 4.1g(i)) and 1,089 of these CD4 Surveillance records were linked to HARS. Similarly, 4,723 HARS records were linked to CD4 Surveillance records but not to SOPHID records (Figure 4.1g(ii)), of which 2,864 linked via CD4 Surveillance to SOPHID. These links were not mutually exclusive and there were 734 common links between HARS, CD4 Surveillance and SOPHID (Figure 4.1g(iii)). However, many of these links were not unique and linked further to multiple HARS or SOPHID records. These links were therefore not considered to be reliable enough for further analysis.

4.3.2 Unique bilateral links and non-linked records

A full understanding of the data and creation of an integrated dataset could only be achieved by consideration of unique bilateral links between two datasets and records that did not link at all. Particular questions that needed answering were: could SOPHID records linked to CD4 Surveillance but not HARS be included in the integrated dataset with the date of diagnosis defined from these data?
Could HARS records linked to CD4 Surveillance but not SOPHID be included in the integrated dataset for analyses that did not require treatment information?

Record linkage provided the potential for any combination of the data to be used but for this thesis, questions of which subset of data to use could only be answered by considering the analyses that were going to be performed on the data and the public health benefits and research knowledge that could be achieved. The analyses planned were:

1. An analysis of CD4 cell counts at the time of HIV diagnosis and the factors associated with late diagnosis (CD4 cell count less than 200 cells/mm^3) and consequent mortality.
2. Quantification of follow-up time for each individual by level of immunosuppression and use of these data to calculate the incidence of first clinical AIDS diagnoses and deaths.
3. Development of algorithms to determine the time of starting ART in order to improve the completeness of these data for further analysis
4. An analysis of the rate of change of CD4 cell counts after starting ART.

Data requirements for the first two analyses were the date of diagnosis, place of diagnosis, demographics (including risk group and ethnicity), date of death and CD4 cell count (only at diagnosis for analysis 1 but all for analysis 2). Country of infection, country of birth and year of arrival were supplementary data required for the analysis of late diagnosis. Data requirements for the third analysis were the date of starting ART, the first date reported to be on ART, date of first clinical AIDS diagnosis and all CD4 cell counts. The last analysis required those for the third analysis and the date of diagnosis, demographics and the place of starting treatment.

I decided that HARS data linked to SOPHID but not to CD4 Surveillance could not substantially contribute to the integrated dataset for this thesis as no dates

were associated with the CD4 cell counts collected in SOPHID. SOPHID data alone was also inadequate for the same reasons. HARS data alone could contribute to analysis 1 where CD4 counts were reported on clinician report forms (CHR) but none of the other analyses. CD4 Surveillance data were not all from confirmed HIV-infected individuals and did not include risk group or ethnicity and therefore could not be included in the integrated dataset without linkage to either HARS or SOPHID. However, these combinations of data could be used for other analyses not considered in this thesis.

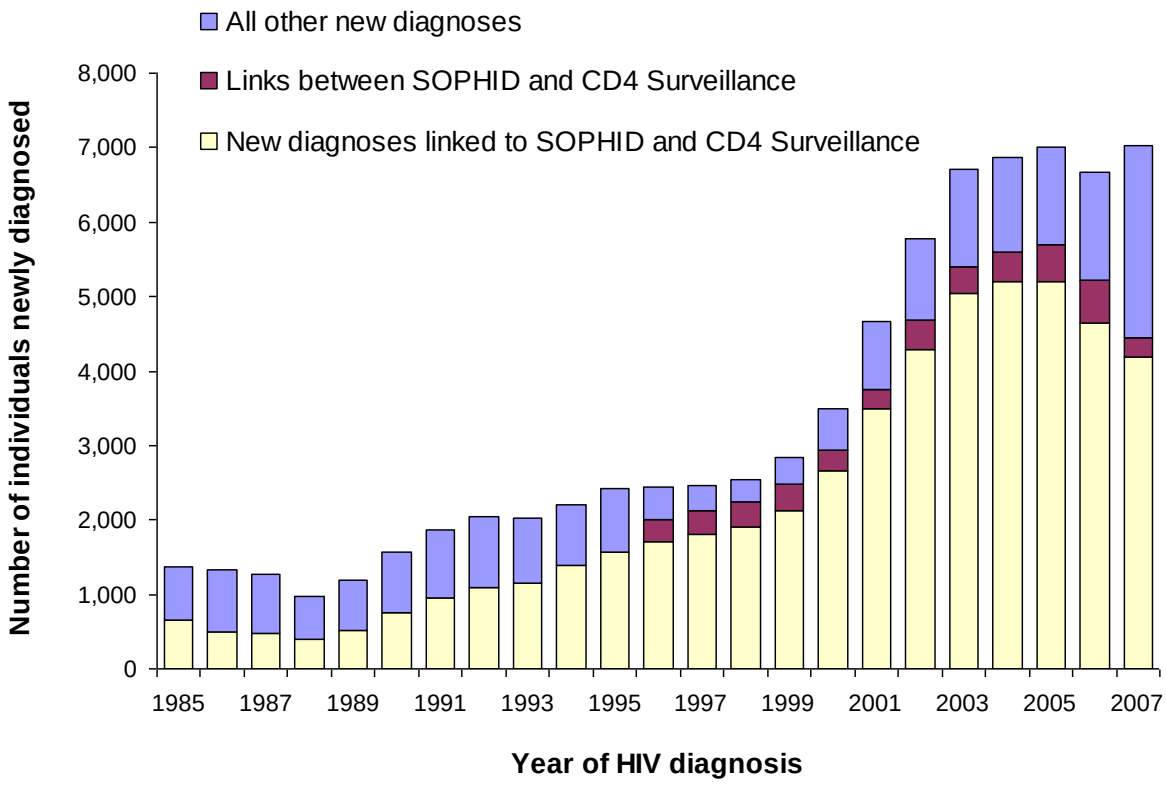
4.3.2.1 Could SOPHID records linked to CD4 Surveillance but not HARS be included in the integrated dataset with the date of diagnosis defined from these data?

Data requested in SOPHID reports included the dates first seen ('datepos') and last seen for care. The date last seen for care in the earliest calendar year in which the individual was reported to SOPHID was derived from the combined dataset. Some records had evidence of previous care reported and many individuals first reported in 1995 were clearly diagnosed prior to that year.

For SOPHID records uniquely linked to CD4 Surveillance but not to HARS, there was potential to use the first CD4 cell count or the date first reported to be seen for care as a proxy for the time of HIV diagnosis. SOPHID records that were not linked to HARS were most likely reported to HARS but with identifiers too different to allow record linkage. Therefore, SOPHID records uniquely linked to CD4 Surveillance but not to HARS could be used to supplement the number of newly diagnosed individuals included in analyses 1 and 2 and the number of individuals starting ART included in analyses 3 and 4. There were 5,204 of these records in total, of which 4,310 were first seen for care after 1995.

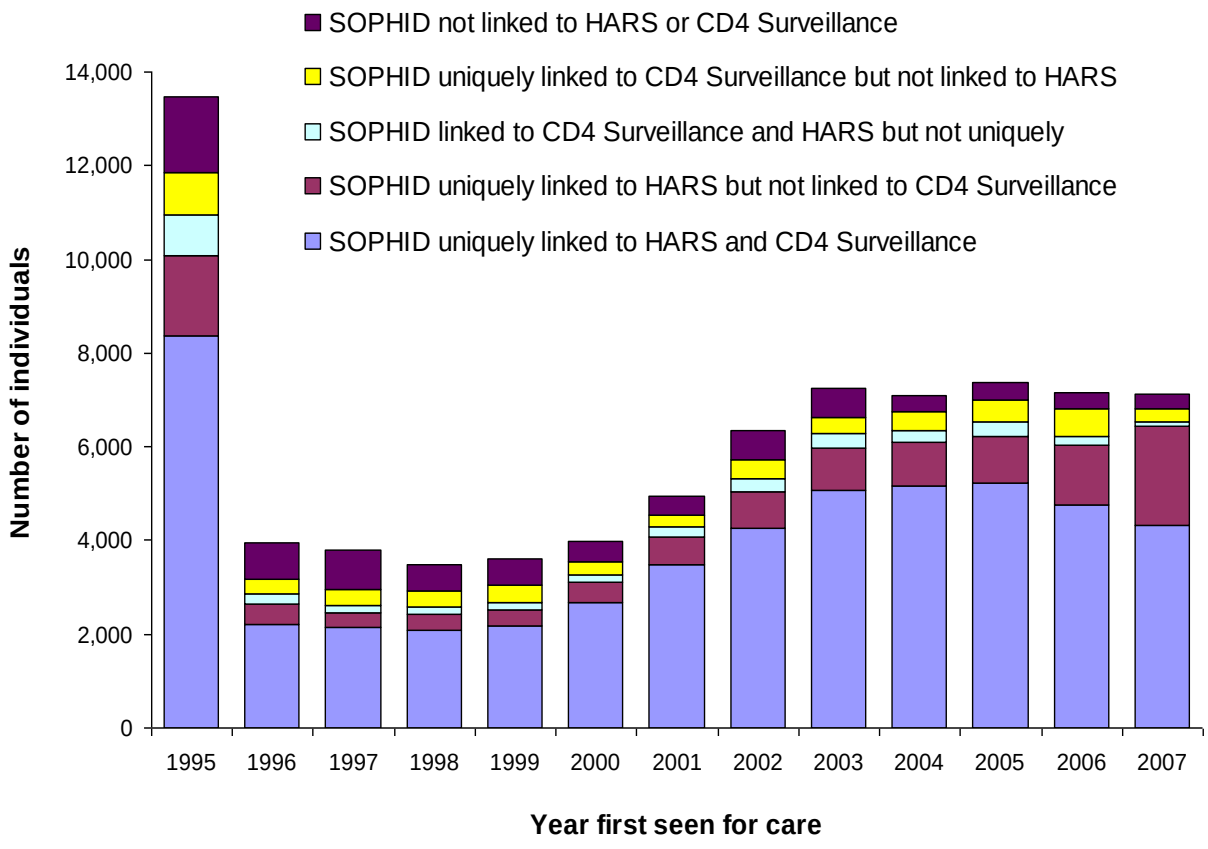
With regards analyses 1 and 2, neither SOPHID nor CD4 Surveillance include data on country of infection or country of birth and therefore these records would have to be excluded from analyses using this information. Secondly, this addition of 4,310 (5.6%) links was thought to be unlikely to markedly increase the proportion of individuals newly diagnosed who could be included in the integrated dataset (Figure 4.3). Finally, if HARS records that were not linked to SOPHID were also included in analyses then individuals could be counted more than once. Therefore, these data were not included in analyses 1 or 2.

Figure 4.3. Overall numbers of HIV diagnoses reported to HARS, numbers linked to both SOPHID and CD4 Surveillance and numbers potentially represented by links between SOPHID and CD4 Surveillance



With regards analyses 3 and 4, linked SOPHID/CD4 Surveillance data already included all necessary variables for these analyses except for the date of first AIDS diagnosis. Additionally, these 4,310 (5.4%) links would not markedly supplement the numbers of individuals first seen for care who could be included in the integrated dataset (Figure 4.4). Therefore, I decided that these data would not be included in the integrated dataset.

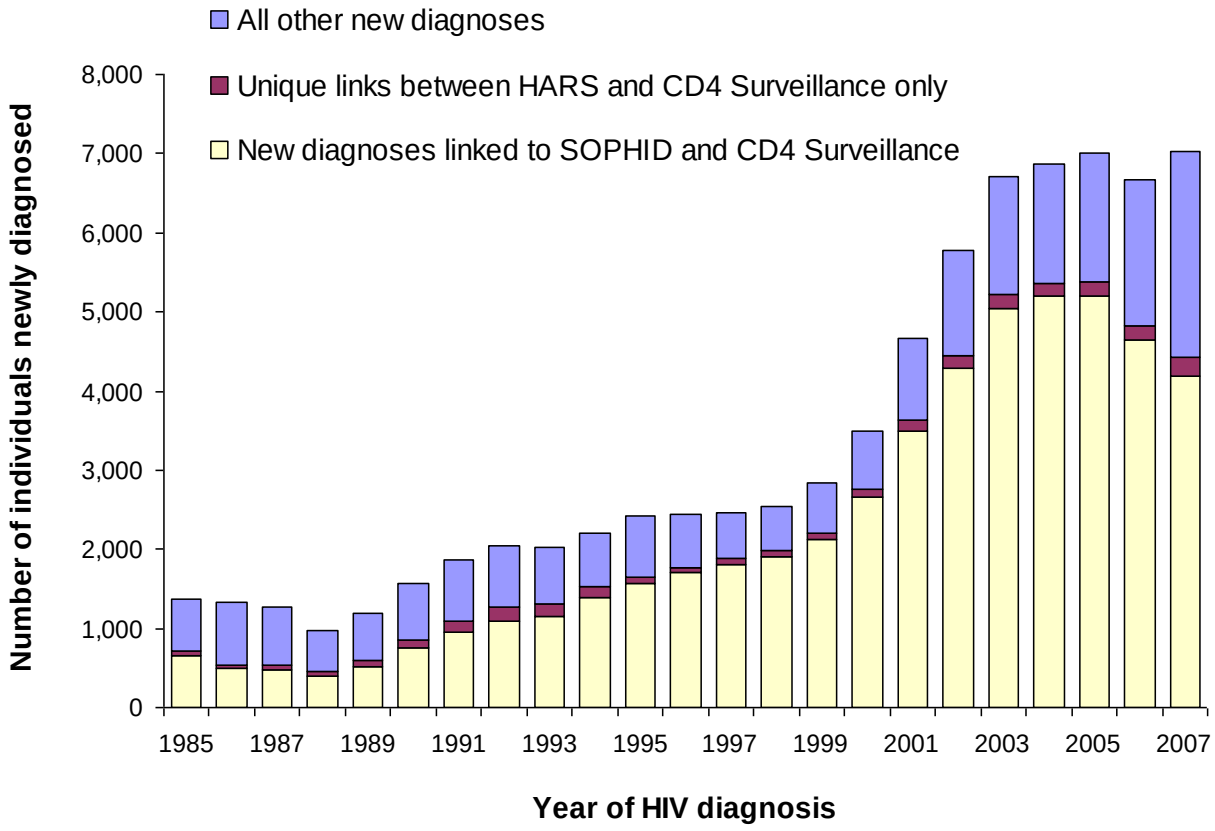
Figure 4.4. Overall numbers of individuals first seen for care reported to SOPHID: numbers linked to HARS and CD4 Surveillance and numbers not linked



4.3.2.2 Could HARS records linked to CD4 Surveillance but not SOPHID be included in the integrated dataset for analyses that did not require treatment information?

HARS records linked to CD4 Surveillance but not SOPHID could be included in the analysis of CD4 cell counts at the time of HIV diagnosis and consequent mortality (analysis 1). However, the additional number (2,754 [3.6%]) of unique links between HARS and CD4 Surveillance that were not linked to SOPHID was relatively small (Figure 4.5). Therefore, these links were not included in the integrated dataset and a consistent dataset of triangulated links between all three datasets was used throughout the subsequent analyses in this thesis.

Figure 4.5. Overall numbers of HIV diagnoses reported to HARS, numbers linked to HARS and CD4 Surveillance only and all other HIV diagnoses



4.4 Integrating data variables in the integrated dataset

4.4.1 Sex, risk group, ethnicity and date of birth

Risk group and ethnicity are reported to HARS and SOPHID whereas sex and date of birth are additionally reported to CD4 Surveillance. Risk group was preferentially assigned according to how it was reported to HARS as it is primarily determined by health advisors at the time of HIV diagnosis or through follow-up by the HPA research nurse. For the 874 (1.7%) individuals for whom this information was not recorded in HARS, 525 (60.0%) could be assigned from SOPHID reports. Sex was also assigned from HARS as risk group and sex are internally consistency-checked within each database to detect records of female MSM, which are then followed-up. Ethnicity was less often missing in SOPHID than in HARS as it is followed-up for SOPHID if missing or reported differently over time. Therefore, ethnicity information was assigned from SOPHID, unless missing, in which case it was assigned from HARS (498/975). As patients may give false dates of birth when taking an HIV test that may get reported to HARS, dates of birth were assigned from SOPHID as these dates, reported from the notes of continuing clinical HIV care, were assumed to be more reliable (1,687 [3.4%] were reported differently to HARS).

4.4.2 Date of first UK diagnosis

The date of HIV diagnosis in the UK is a vital piece of information in HIV surveillance systems and in the integrated dataset. Data about the date of HIV diagnosis in the UK are reported to HARS (the date of HIV diagnosis and date of first positive test), SOPHID (the date first seen for clinical care [earliest of the dates last seen for care in each survey period], date of first positive test, date of

AIDS diagnosis and date of starting ART), and CD4 Surveillance (the date of first CD4 cell count measurement). In HARS, the date of HIV diagnosis attributed to a patient is the first of any date reported except for the date of previous positive test because this may be abroad. In the integrated dataset, dates from other surveillance systems were used to investigate whether the date of previous positive test reported to HARS could be validated. Dates from SOPHID and CD4 Surveillance, which occurred before the date of HIV diagnosis in HARS, were also examined to see whether they were likely to have occurred in the UK. Validated dates were used to update the date of HIV diagnosis in the integrated dataset according to the principle of assigning the first date reported as employed in HARS.

There were 427 records in the linked HARS dataset that had a previous positive test in the UK in a year before the date of HIV diagnosis. A further 908 had a reported previous positive test abroad in a year prior to diagnosis in the UK. Almost two-thirds, 63.9% (273) had a date first seen for clinical care in the same calendar year as the date of HIV diagnosis and 229 of these also had the first CD4 cell count in the same calendar year as the date of HIV diagnosis. Therefore, the majority of the links between HARS, SOPHID and CD4 Surveillance did not validate the date of previous positive test as being in the UK. Therefore, this analysis supports the internal HARS protocol of not using earlier dates of first positive tests to update the date of HIV diagnosis.

There were 2,172 individuals whose earliest date seen for care reported to SOPHID was before the date of HIV diagnosis reported to HARS: for 1,010 it was within 1 year, 535 within 2 years and 627 more than 2 years (Table 4.1).

Table 4.1. Number of years between the date first seen for care reported to SOPHID and the date of HIV diagnosis reported to HARS where the date first seen for care was earlier than the date of HIV diagnosis (n=2,172)

Number of years between the date first seen for care and the date of HIV diagnosis	0	1	2	3	4	5	6	7	8	9	10	11	12
Number of individuals	1,010	535	156	108	86	62	53	41	37	30	24	15	15
(%)	46.5	24.6	7.2	5.0	4.0	2.9	2.4	1.9	1.7	1.4	1.1	0.7	0.7

Where SOPHID recorded an earlier calendar year of diagnosis than HARS (n=1,162) and the first year seen for care was only one year earlier than the year of diagnosis (n=535), 54.6% of records had their first CD4 count in the year of diagnosis and 39.5% in earlier years. Where the first year seen for care was more than one year earlier than the year of diagnosis (n=627), 18.3% of records had their first CD4 count in the year of diagnosis and 79.0% in earlier years. This suggested that the date first seen reported to SOPHID was valid even if it was earlier than the date of HIV diagnosis reported to HARS and therefore it was used to update the date of HIV diagnosis in the integrated dataset.

There were 4,978 individuals (9.6% of 51,969 records in the integrated dataset) who were reported to SOPHID as having had their first HIV diagnosis (datepos) in a calendar year before their date of diagnosis reported to HARS or the earliest date last seen as reported to SOPHID. Records for the majority of these individuals (71.3% [3,569]) had their first CD4 cell count in the same year as the date of diagnosis/earliest date last seen. This indicates that these first dates of HIV diagnosis as reported to SOPHID (datepos) should not be used to update

the date of HIV diagnosis in the integrated dataset because many of these dates are likely to be prior to arrival in UK. However, these dates could be used as probable evidence of prior diagnosis abroad and be used to exclude records from subsequent analyses.

In contrast, there were 10,452 records where the SOPHID datepos was earlier in the same calendar year as the year of diagnosis/earliest year seen for care. Of these, 4,246 were not reported to be the 1st or 15th of the month, or the 30th June (likely to reflect retrospective approximations). These were considered reliable enough to update the date of HIV diagnosis in the integrated dataset.

There were 1,052 individuals (2.0% of 51,969 linked between HARS and SOPHID) who were reported to have started ART in a calendar year before their date of diagnosis/earliest date last seen. Most (864 [82.1%]) of these individuals had their first CD4 cell count measured in the same year as the year of diagnosis/earliest year seen for care. However, almost all (1,036 [98.5%]) had a reported date of diagnosis reported to SOPHID (datepos) and most of these (719 [69.4%]) were earlier or the same as the reported date of starting ART. This suggests that many of the dates of starting ART were correct but that they were likely to be prior to arrival in the UK. Of the 1,052 individuals, there were 598 (56.8%) who had been reported by a clinician of whom 309 (51.7%) had a reported year of arrival in the UK. A large percentage of these individuals (74.4% [230]) were reported to have arrived in the UK after they started ART. These analyses indicate that the reported date of starting ART should not be used to update the date of HIV diagnosis in the integrated dataset. However, these dates could be used as probable evidence of prior diagnosis and treatment abroad and be used to exclude records from subsequent analyses.

The majority of individuals in the integrated dataset (73.8% [38,345]) had their first CD4 cell count in the same year as their date of diagnosis/earliest date last seen (82.4% [34,629/42,007] between 1996 and 2007). Only a small percentage (1.8% [949]) had their first CD4 cell count before the year of their first HIV diagnosis/year first seen for care. Manual review of a subset of the latter showed that most were likely to be due to inappropriate deduplication in CD4 Surveillance and therefore, this small proportion of records was excluded from further analysis. However, the first CD4 cell count was in the same calendar year but earlier than the date of diagnosis/earliest date last seen for 2,061 records. There was good evidence from manual review that these CD4 cell counts were consistent for each individual and often reported from the same laboratory. Therefore, the earliest date of these CD4 cell count measurements was used to update the date of HIV diagnosis in the integrated dataset.

In summary, in the integrated dataset, the date of HIV diagnosis in E,W&NI was assigned to be the earliest of the following: i) the earliest date of HIV/AIDS diagnosis or death as reported to HARS; ii) the earliest date last seen for care as reported to SOPHID; iii) the earliest date of HIV diagnosis (datepos) reported to SOPHID if it was in the same year as one of the above; iv) the date of the first CD4 cell count reported to CD4 Surveillance if it was in the same year as one of the above. The dates of first positive test from HARS and SOPHID (if prior to the year of diagnosis already defined) and the date of starting ART from SOPHID were not valid for updating the date of HIV diagnosis in the integrated dataset because of the likelihood that many of these dates occurred prior to arrival in the UK. This algorithm is consistent with the internal HARS protocol for processing multiple reports from the same individual that include earlier dates.

4.4.3 Date of first AIDS diagnosis

There were 9,782 records in the integrated dataset for individuals who had been reported to HARS as clinically diagnosed with AIDS. The earliest date of AIDS diagnosis reported to SOPHID was earlier than the date reported to HARS for 1,205 individuals. Two-thirds (67.2% [810]) of these were in the same calendar year, a fifth (21.2% [255]) were in the previous calendar year and 2.1% (25) were in a calendar year more than six years previously. These dates reported to SOPHID were used to update the earliest date of AIDS diagnosis in the integrated dataset. There were only two records with a clinical stage of AIDS first reported earlier than both the dates of AIDS reported to HARS and SOPHID. These were also updated in the integrated dataset.

There were 7,446 records in the integrated dataset without a date of AIDS reported to HARS but with AIDS reported to SOPHID. Almost all (99.4% [7,398]) of these had both a date of AIDS diagnosis and a clinical stage of AIDS reported to SOPHID. The date of AIDS diagnosis was the same as, or earlier than, the earliest date last seen with clinical stage of AIDS for almost all (99.8% [7,382]) of these records. There were 519 (7.0%) records with the earliest date of AIDS diagnosis reported to SOPHID earlier than the combined date of HIV diagnosis in the integrated dataset. Where the date of AIDS diagnosis reported to SOPHID was in the same calendar year as the combined date of HIV diagnosis (62.0% [322]), two-thirds (66.5% [214]) had the same month and three-quarters (77.6% [250]) had the AIDS diagnosis reported as the 1st or 15th of the month. For these records, the combined date of HIV diagnosis in the integrated dataset was not updated but the date of AIDS diagnosis was updated to be the same as the date of HIV diagnosis. Where the date reported to

SOPHID was in an earlier calendar year than the combined date of HIV diagnosis (38.0% [197]) and where year of arrival in the UK was reported to HARS (14.6% [76]), the majority (68.4% [52]) had AIDS prior to arrival in the UK. Therefore, these dates of AIDS diagnosis were used to update the date of AIDS diagnosis but not the date of HIV diagnosis in the integrated dataset and assumed to occur before arrival in the UK. There were 6,927 remaining records with a date of AIDS diagnosis reported to SOPHID but not to HARS. All of these dates were used to update the date of AIDS diagnosis in the integrated dataset.

4.4.4 Date of last contact with services/date of death

The last date of contact with HIV services was defined in the integrated dataset as the date of latest CD4 cell count, the date of the latest report to SOPHID or the date of the latest AIDS diagnoses prior to death. The majority (85.7% [44,579]) of these dates in the integrated dataset were in the same calendar year. There were 32,658 (62.8%) with the date last seen reported to SOPHID after the latest CD4 cell count measurement reported, 11,919 (22.9%) where these dates were the same, and 7,392 (14.2%) where the latest CD4 cell count was after the date last seen reported to SOPHID. Where the date last seen was after the latest CD4 cell count, 83.9% (27,388) were in the same year, 10.1% (3,313) in the following year, and 0.8% (277) more than five years later. Where the latest CD4 cell count was after the date last seen, 71.3% (5,272) were in the same year, 21.1% (1,559) in the following year, and 2.2% (159) more than five years later. There were only 22 records where the last date of contact with HIV services was updated with the first AIDS diagnosis because this was prior to death (indicating contact with services) but after the latest CD4 cell count and the latest report to SOPHID.

The date of death was defined primarily from HARS as this was the traditional basis of surveillance of deaths among HIV-infected individuals and includes record linkage between HARS and ONS (Office of National Statistics data) and follow-up (Section 2.2). Dates of death were first supplemented from SOPHID and then from CD4 Surveillance. There were 4,415, 382 and 38 records in the integrated dataset for individuals reported to have died with dates of death derived from HARS, SOPHID, and CD4 Surveillance respectively. However, 10 (0.2%), 24 (6.3%) and 27 (71.1%) of these records were excluded from subsequent analysis as there were either reports of CD4 cell counts measurements or contact with HIV services after the date of death. These records could not be resolved without further follow-up because the discordant dates could arise from either coding errors in any of the dates reported, incorrect deduplication of records, or incorrect record linkage between any of the HIV surveillance databases and the ONS database of deaths. There were a final total of 4,774 deaths between 1995 and 2007 in the integrated dataset.

4.4.5 Previous diagnosis before HIV diagnosis in the UK

There were 1,304 records in the integrated dataset with a date of a previous positive HIV test reported to HARS in an earlier calendar year than the date of HIV diagnosis (earliest from all sources). Three-quarters (74.5% [971]) indicated that the previous HIV test occurred abroad. Most of these dates were approximations (e.g. 30/06/2001) and therefore only the calendar year was considered. Additionally, there were 936 records with a comment on a HARS form indicating a previous positive HIV test but no reported date.

There were 4,844 records in the integrated dataset that had an HIV diagnosis (datepos) reported to SOPHID in a calendar year before the date of HIV

diagnosis. A fifth of these (19.0% [920]) also had evidence of a previous positive HIV test reported to HARS. A further 267 and 1,011 records had a date of an AIDS diagnosis and date of starting ART, respectively, reported to SOPHID in a calendar year before the date of HIV diagnosis in the integrated dataset, of which 89 and 191 had not been previously identified.

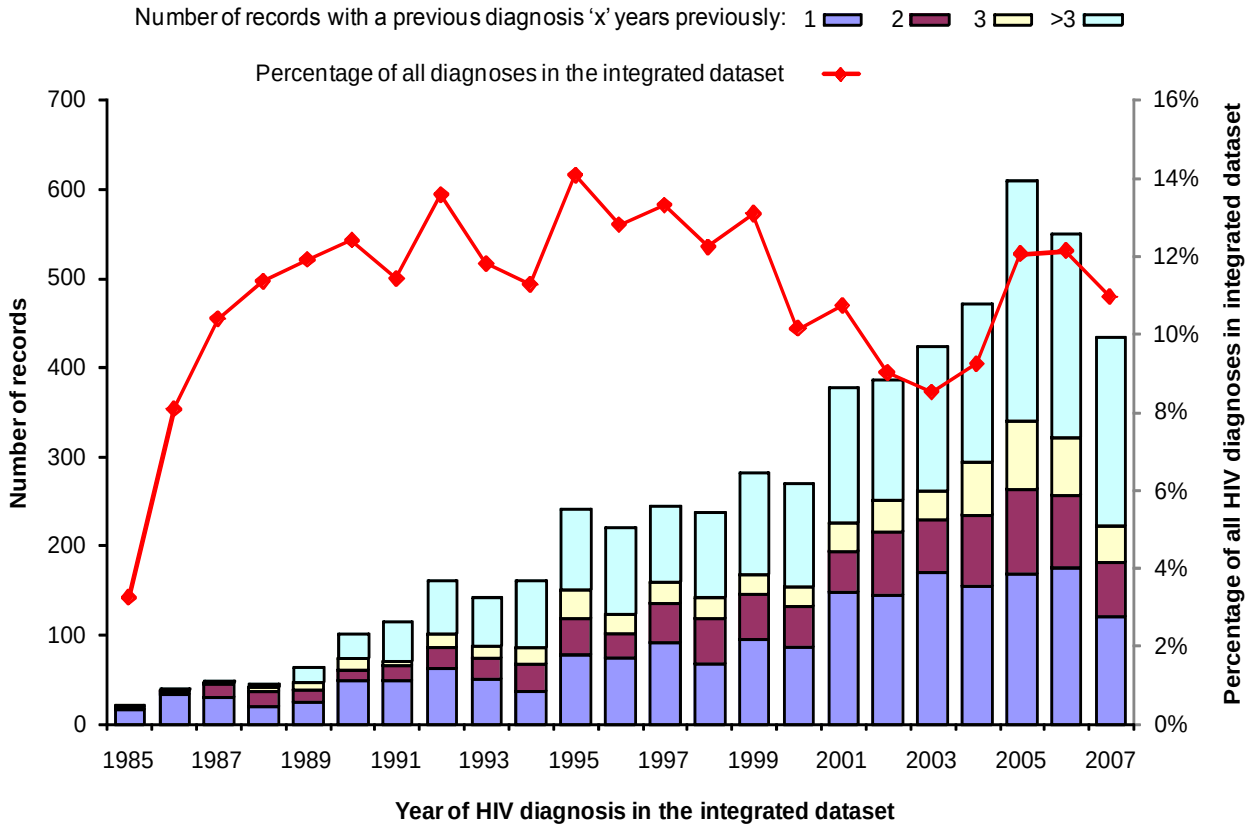
In total, there were 6,444 (12.4%) records in the integrated dataset with evidence of a previous HIV diagnosis, of which 5,766 had a reported year of previous HIV diagnosis. The year of previous HIV diagnosis was one year earlier than the date of HIV diagnosis in the integrated dataset for a third of records (33.9% [1,955]) but more than five years earlier for a quarter of records (24.9% [1,438]) (Table 4.2).

Table 4.2. Number of calendar years between HIV diagnosis in the integrated dataset and the reported year of previous HIV diagnosis (where reported)

Number of years between diagnosis in the integrated dataset and previous HIV diagnosis	1	2	3	4	5	6	7	8	9	10	11-20	>20
Number of individuals	1,955	937	584	483	369	299	237	200	139	143	399	21
(%)	33.9	16.3	10.1	8.4	6.4	5.2	4.1	3.5	2.4	2.5	6.9	0.4

By year of HIV diagnosis in the UK in the integrated dataset, there was an increasing trend in the number of records with a previous diagnosis (Figure 4.6). However, since 1987, this number accounted for a fairly stable percentage of all HIV diagnoses in the integrated dataset, with an annual mean of 11.5%.

Figure 4.6. Number of records with a previous HIV diagnosis and the annual total of HIV diagnoses as a percentage of all HIV diagnoses

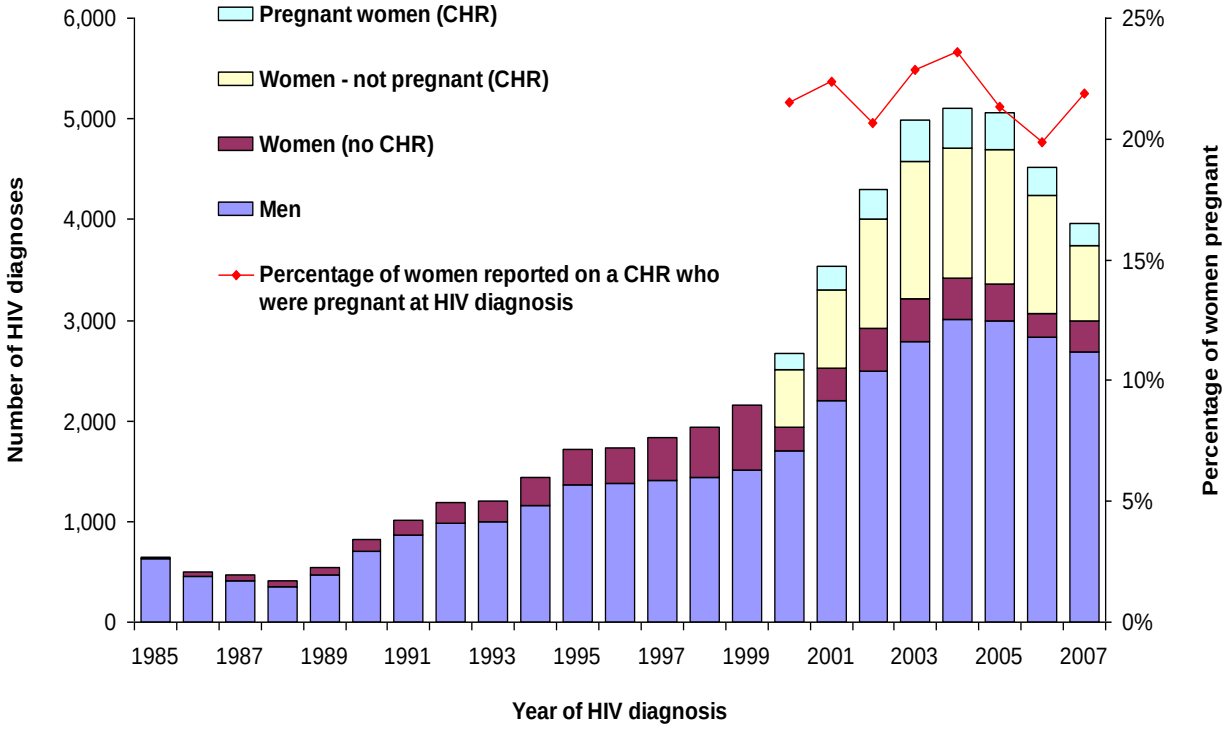


4.4.6 Pregnancy and recent infection status at HIV diagnosis

There were 3,657 women from E,W&NI who were identified in HARS as pregnant at HIV diagnosis with an HIV-related event within 91 days of their date of HIV diagnosis reported to HARS. These records were either reported from an antenatal clinic with pregnancy marked on the form or with pregnancy or antenatal testing marked as the reason for test. The integrated dataset included records for 2,487 (68.0%) of these pregnant women but only 2,429 (97.7%) were updated as being pregnant at HIV diagnosis because the event date reported on the HARS form at the time of pregnancy was within nine months of their date of HIV diagnosis in the integrated dataset (earliest from all sources).

Pregnancy at the time of HIV diagnosis has only been collected in HARS since 2000 on the CHR (Figure 4.7). These reports accounted for an annual mean of 17.2% of all women diagnosed with HIV infection in the integrated dataset between 2000 and 2007 but an annual mean of 21.8% of all women reported on a CHR between 2000 and 2007. HIV diagnoses among women reported from laboratories without a CHR cannot be classified by pregnancy status.

Figure 4.7. Number of HIV diagnoses in the integrated dataset by sex, showing whether women were reported on a CHR and pregnant at the time of diagnosis.



There were 2,998 records in the integrated dataset with some evidence of recent infection reported to HARS. The variables used were symptoms of seroconversion indicated on the CHR, or in the comment field on any form, or a negative HIV test within a year of the date of HIV diagnosis. The majority (60.6% [1,817]) were identified by a previous negative HIV test, with 31.1% (932) identified by symptoms of seroconversion and/or 28.4% (851) identified by a comment (Figure 4.8). The introduction of the CHR and increases in HIV testing (identification of previous negative HIV tests) increased the percentage of all HIV diagnoses in the integrated dataset with evidence of recent infection from 2% or less before 2000 (except 1998) to over 10% in 2005 and 2006.

Figure 4.8. Number of HIV diagnoses with evidence of recent infection in the integrated dataset according to how this evidence was reported.

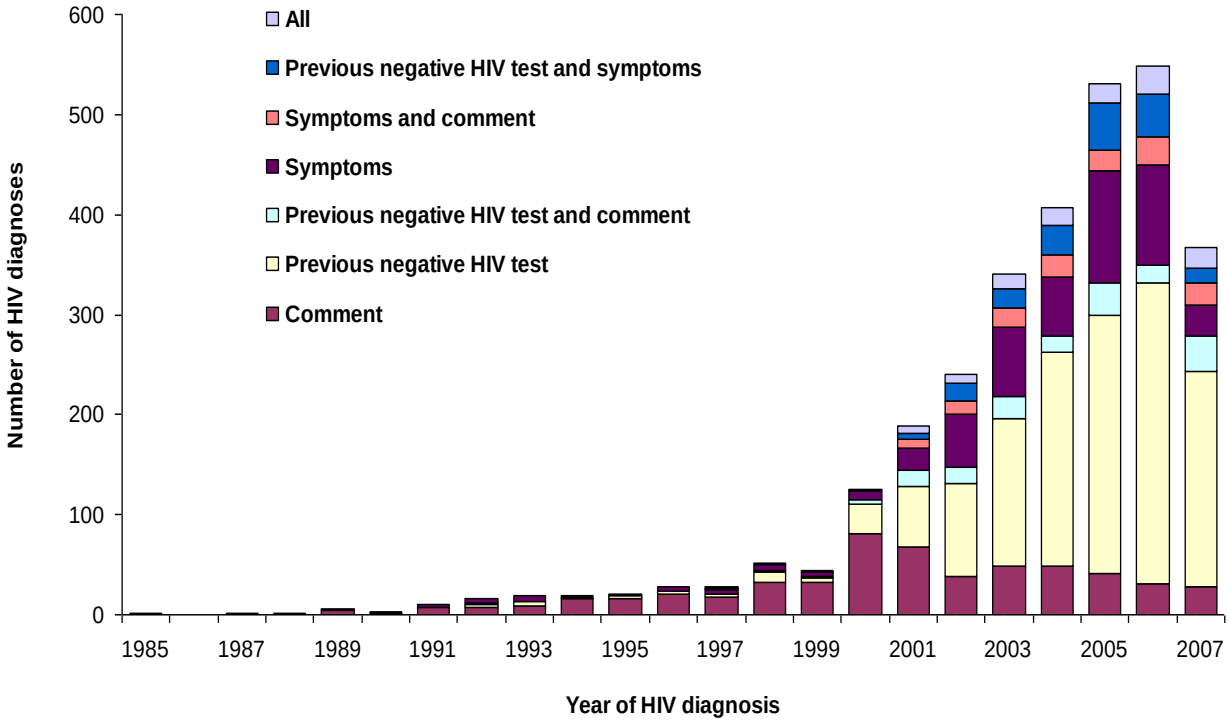
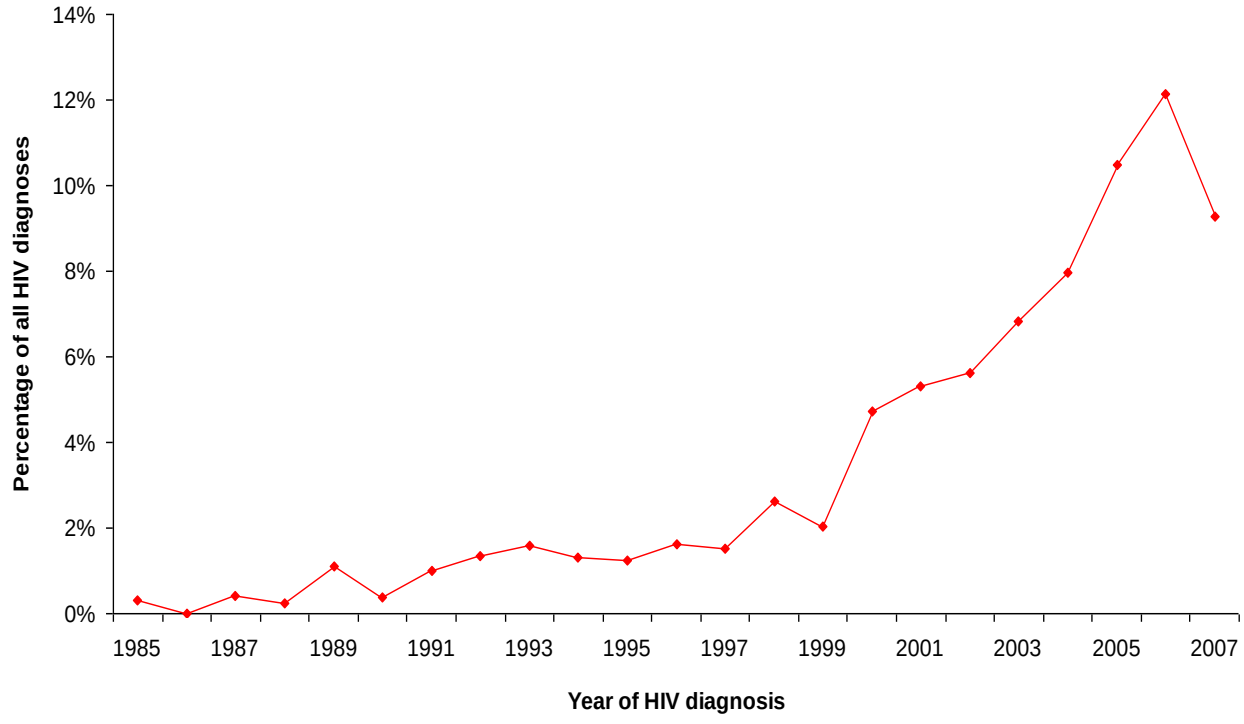


Figure 4.9. Percentage of all HIV diagnoses in the integrated dataset with evidence of recent infection.



4.4.7 CD4 cell counts

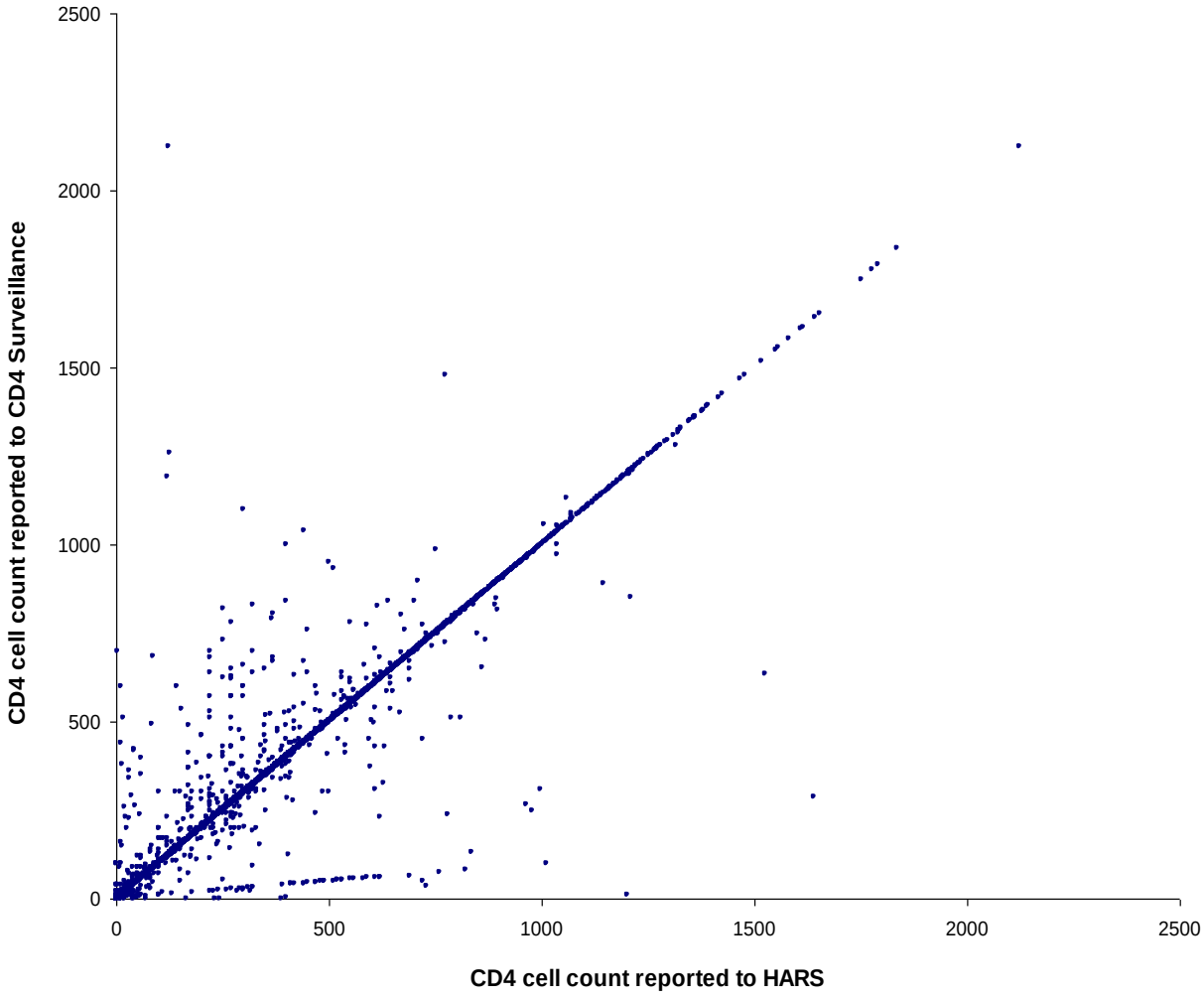
CD4 cell counts were available from CD4 Surveillance, HARS and SOPHID. If reporting were complete: CD4 Surveillance would include all CD4 cell counts for an individual over time; HARS would include CD4 cell counts at the time of first presentation in the UK for the two-thirds of presentations that were reported by clinicians since 2000; and SOPHID would include the last CD4 cell count during the survey period for all individuals since 2000 (Chapter 2). As the date of CD4 cell count was required for subsequent analyses but not collected in SOPHID, only CD4 cell counts at the time of first presentation from HARS were used to supplement CD4 Surveillance.

4.4.7.1 Identification and deletion of duplicate CD4 cell counts

There were 1,242,903 CD4 cell counts in the CD4 Surveillance dataset and 28,263 CD4 cell counts in the HARS dataset to the end of 2007. A total of 902,838 (72.6%) and 21,100 (74.7%) of these CD4 cell counts were from the 50,953 individuals in the integrated dataset. Integration led to the creation of 12,397 duplicate CD4 count measurements on the same date. As expected, there was a high correlation (r=0.97) between the duplicate measurements but 1,926 (0.21% of all measurements) CD4 cell counts were not the same (Figure 10). Possible explanations are that some samples were analysed twice on the same day for quality control, results were rounded to the nearest ten before being reported or that clerical errors occurred. To maintain maximum consistency of the CD4 cell counts in the integrated dataset, these duplicate measurements reported to HARS were deleted. A further 6,187 of the remaining CD4 cell counts reported to HARS were deleted because the preceding or following CD4 cell count reported to CD4 Surveillance was the same. Finally,

2,145 CD4 cell counts from HARS were deleted because of the following: they were measured within 14 days of a CD4 cell count reported to CD4 Surveillance; they were not the first CD4 cell count reported; they were earlier than the date of HIV diagnosis in the integrated dataset and appeared to be mostly spurious dates; they had CD4 cell counts of zero that did not appear to match subsequent CD4 cell counts; they were multiples of ten of subsequent CD4 cell counts. There were 3,203 CD4 cell counts from HARS remaining in the integrated dataset.

Figure 4.10. Correlation between duplicate CD4 cell count measurements reported both to CD4 Surveillance and HARS.



Duplicate measurements were also identified within CD4 Surveillance records. There were 16,605 records with identical CD4 cell counts reported within 14 days of a previous measurement. Three quarters (75.7% [12,907]) were from the same laboratory and the majority (85.0% [10,759]) of these were from two large London laboratories with multiple clinics on different sites. For one laboratory, 96.4% (8,329) of these duplicates occurred in 2004 and 2005 with 93.3% (7,821) within 2 days of each other. In the other laboratory, 82.4% (1,795) occurred in 2006 with 91.3% (1,651) within 1 day of each other. These duplicates were therefore deleted along with a further 404 repeated measurements at other laboratories identified in this manner. The majority of repeat measurements from different laboratories (96.0% [3,569/3,718]) were from four laboratories. These tended to occur within two days of the previous measurement and in specific years and therefore 3,471 CD4 cell counts were also deleted as probable duplicates.

4.4.7.2 Identification and deletion of invalid CD4 cell counts

Potentially incorrect measurements are often validated in clinical practice by testing a second sample shortly afterwards[233;234]. There were 35,189 (4.0%) CD4 counts in the integrated dataset that were followed by a repeat measurement within 14 days for 17,067 individuals. For 5,268 (30.9%) individuals, their first CD4 cell count was repeated within 14 days with a median change in CD4 cell counts of 1 (inter-quartile range [IQR] -22, 40; range -1,240, 1,165) cells/mm^3. Without knowing which of these was the most accurate reflection of the true CD4 cell count, the measurement bias was minimised by averaging the values, using the first date and deleting the 5,268 repeat measurements.

There were 12,782 individuals with 27,518 CD4 cell counts followed by a repeat measurement within 14 days and also preceded by a measurement within 182 days. CD4 cell counts that were likely to be incorrect were identified as being either substantially higher, or lower, than both the preceding and repeat measurements to exclude expected variation. The median change in CD4 cell counts prior to the suspect measurement was 42 (IQR 21, 80; range -1,655, 1,995) cells/mm^3 and the median change to the repeat measurement was 3 (IQR -30, 54; range -1,900, 1,690) cells/mm^3. There were 2,807 (22.0%) individuals with 3,374 (12.2%) CD4 cell counts that were more than 100 cells/mm^3 (arbitrarily chosen) higher or lower than both the preceding and repeat measurements. Manual examination of the longitudinal CD4 cell counts for a number of these individuals showed that there was a mixture of erroneous single measurements and whole histories of highly variable measurements. The single erroneous measurements should be excluded from the former, whereas the latter should be further investigated to ensure that the CD4 cell counts and dates reliably represented the measurements for a single individual. However, it was not possible to distinguish between the two situations without either manual review or much more complicated statistical analysis, which was not appropriate at this stage. Substantial bias could be introduced by including or excluding all records for individuals who had highly variable CD4 cell counts depending on whether these records incorrectly represented more than one individual due to unspecific deduplication or not. Conversely, bias could be slightly reduced by excluding single erroneous CD4 cell counts that were likely to be due to recording errors and the removal of a single CD4 cell count from individuals who had highly variable CD4 cell counts was unlikely to create significant bias. On consideration of these potential impacts on the dataset, and the difficulty in

identifying whether the repeat CD4 cell count or the measurement before it were erroneous, I averaged the potentially erroneous and repeat CD4 cell counts, used the first measurement date and deleted the 3,374 repeat measurements.

There were a total of 881,415 CD4 cell counts for the 50,953 individuals in the final integrated dataset.

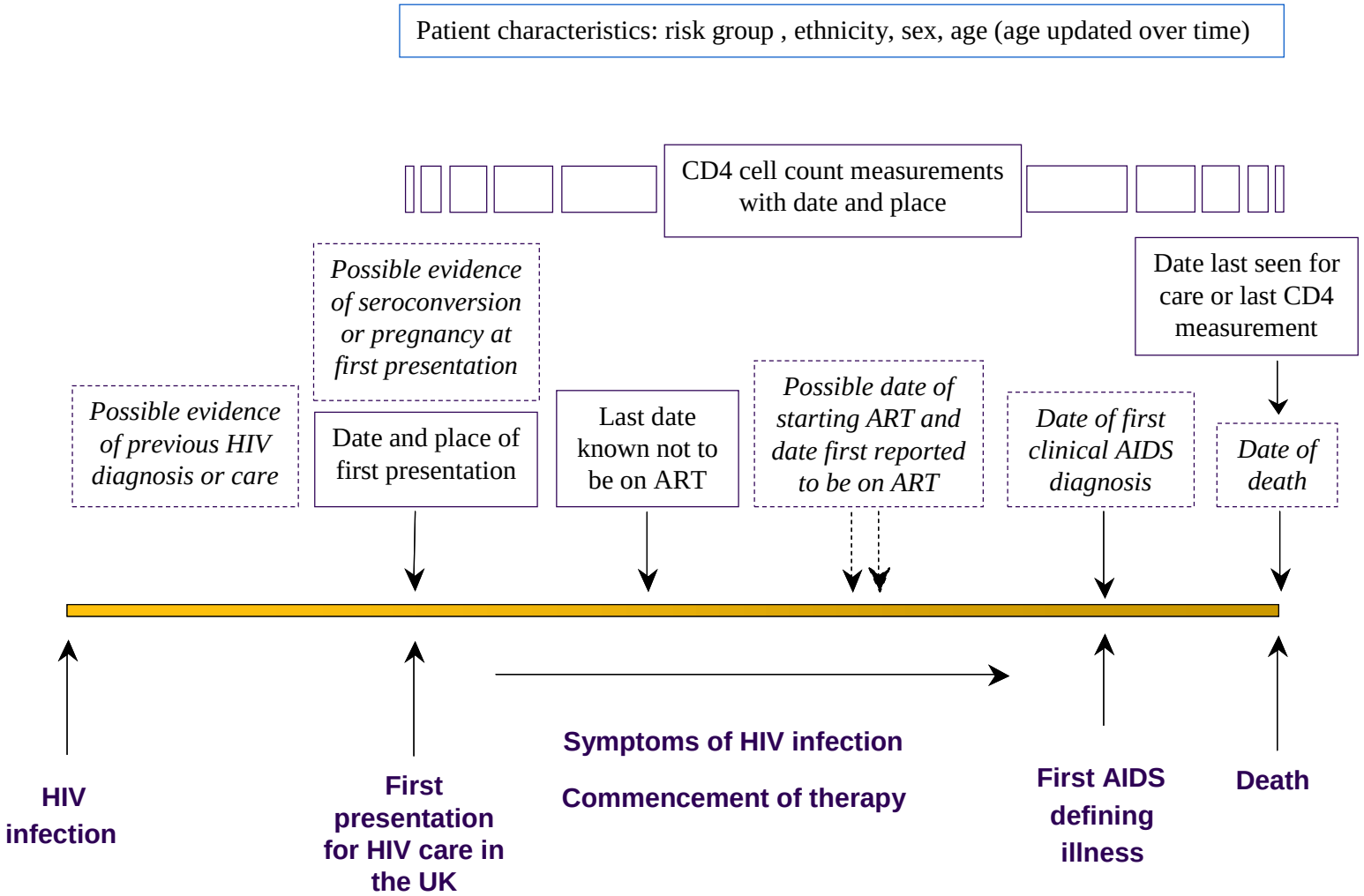
4.5 Conclusion

This chapter shows how a single integrated dataset with coherent patient records was created by the triangulation of records and the justifiable selection and rationalisation of their constituent variables. The process also reduced missing information in each dataset. Numbers remaining for analysis were still equivalent to multinational cohort collaborations (such as EUROSIDA[235]) while having the benefits of national representativeness, good ethnicity information and complete patient follow-up from first diagnosis in the UK.

This was the first time that an integrated dataset had been created from these three national HIV surveillance systems to create a single coherent sequence of events for each individual (Figure 4.11). Patients in this dataset can now be followed from the first known diagnosis of HIV infection or first presentation for HIV care in the UK, through first development of AIDS defining illness and initiation of ART to death. Their CD4 cell counts in relation to this process can be monitored. The effects of patient factors such as risk group, ethnicity, pregnancy status, geographical region and age can be analysed with the latter three specific to time periods along the process (Figure 4.11).

Triangulation supported and validated record linkage as well as enabling record linkage between two datasets via the third dataset. It revealed further inappropriate deduplication within CD4 Surveillance as previously identified through record linkage. Records were excluded because of multiple links or inconsistent information. These validations reduced the number of records in the integrated dataset but increased confidence in the reliability of the data. Resources may need to be redirected away from follow-up for missing information to follow-up to resolve inconsistent information. A reasonable and useful assumption made was that record linkage between HARS and SOPHID could provide a solid basis for analysis and further record linkage. Other record linkage that did not result in reliable triangulation was not included in the integrated dataset largely because certain information was missing or because it would prevent a robust analysis of bias.

Figure 4.11. Timeline of monitored events for a single individual with diagnosed HIV infection.



Integration from multiple sources also provided insight about data quality that could not be determined or quantified from single surveillance systems. For example, there was evidence of substantial and fairly stable proportions previously diagnosed before first presentation in the UK as reported to HARS (around one individual among every eight newly diagnosed). These data for Africans could be compared to the results of Project Nasah, which showed that just over a fifth of African respondents living in the UK were probably first diagnosed with HIV prior to their residence in the UK[236]. Similar findings have also been reported from the United States of America when surveillance data were compared to an interview survey[237]. Previous diagnosis may be important to consider in subsequent analyses and may indicate some under-reporting or an inability to link records within HARS. There was also a clear indication of under-reporting of AIDS to HARS or an inability to link these records in HARS. Integrated data about AIDS diagnoses was crucial for the reliable quantification of AIDS incidence (Chapter 7). In summary, integration allowed sufficient resolution of this information for further analysis but raised questions for further investigation.

The process of integration also provided insights to data from HARS as a single source. There was an annual mean of 17.2% of all women diagnosed with HIV infection who were pregnant between 2000 and 2007 compared to 21.8% of women reported on a CHR. This suggests that a substantial number of women were being diagnosed through antenatal screening but may not be recorded as such if a CHR is not received. This provides further justification for the need to investigate data from surveillance (including linking to other datasets such as

that of the National Study of HIV in Pregnancy and Childhood) and to adjust it for missing information or for known biases in data collection but also suggests that further improvements in surveillance systems could prevent these discrepancies. This is particularly important because of the high public health policy impact of surveillance data[238]. Furthermore, these investigations also indicated that detection of seroconversion at the time of HIV diagnosis may be increasing. This is an interesting finding with significant public health ramifications because it may provide further evidence of increased HIV testing rates in E,W&NI[239;240]. Earlier diagnoses suggest some success of HIV testing policies and/or increased health seeking behaviour[111;241].

Integration did reveal limitations of the surveillance datasets other than under-reporting. In particular, records from CD4 Surveillance that had been inappropriately deduplicated had to be excluded and it is possible that some of these were not identified and remain in the dataset. However, it would be difficult to identify these without substantial manual review and follow-up with clinics. Furthermore, the resulting misclassification bias was likely to result in non-detection of differences rather than spurious results and the effect was likely to be limited by the large number of records available for analysis. Additionally, duplicate and repeat CD4 cell count measurements were identified in CD4 Surveillance. Such issues are expected in observational cohort data because of laboratory logistics (tests contracted out may be recorded in two places), open access to GUM services (testing at diagnosis and retesting elsewhere on entry to care)[3], and the natural and measurement variability of CD4 cell counts[31;43;242-244]. These were accounted for by averaging repeat CD4

cell counts and deleting duplicate CD4 cell counts but this was unlikely to fully eradicate measurement bias and it is likely that not all duplicate and repeat CD4 cell counts were identified. This measurement bias must be considered in subsequent analyses.

The issues of data quality raised above may affect surveillance data more than research data due to the integration of surveillance data from a larger number and variety of sources and the potential for more intensive and expensive efforts to clean smaller research datasets. However, as cohorts expand, more research on the algorithms for data validation and processing may be required.

In conclusion, record linkage and triangulation can be used to validate data but also indicate that variables are duplicated in different surveillance systems. Integration of data at national level has the potential to reduce and rationalise data collection and the burden on data providers but record linkage must be high to achieve completeness. However, if these data are all captured routinely at the service delivery level, and not specifically for surveillance needs, then improvements will not reduce data collection at that level (may apply to CD4 Surveillance and SOPHID where electronic data exist locally, but not to the completion of paper-based HARS forms). At the national level, some resources are likely to need redirection from follow-up of individual records to data validation and confirmation of record linkages. This may be achieved most effectively if record linkage and validation are ongoing as reports are received. These would be steps towards full integration of these surveillance systems.

Chapter 5. Potential for bias in analyses of the integrated dataset

5.1 Introduction

This chapter scrutinises expected sources of potential bias to facilitate the clear interpretation of subsequent analyses of the integrated dataset as recommended for observational studies[245]. Differences in the integrated dataset not solely related to the outcome under analysis could otherwise cause spurious conclusions to be drawn. The key types of bias investigated in this chapter were unequal inclusion of records in the integrated dataset, differential CD4 cell count monitoring, disparate loss to follow-up (LTFU) and unequal inclusion of events due to the varied proximity of CD4 cell counts. Inclusion bias was addressed in detail because it can substantially affect all epidemiological studies but differential monitoring and LTFU were addressed specifically as these types of bias often affect cohort analyses[246]. Observational cohorts based on clinic populations may be particularly sensitive to bias introduced through left, right and interval censoring because patient attendance may be related to their clinical status[247-249]. Analyses of incidence (Chapter 7) and of rates of change in CD4 cell counts (Chapter 9) were statistically approached as cohort analyses.

Bilateral record linkage has been shown to find matches differentially between groups of individuals (Section 3.5.1.6). Creation of the integrated dataset also involved further exclusion of records such that those included had different characteristics to those excluded. A statistical analysis of the factors associated with inclusion in the final integrated dataset was therefore undertaken to characterise the potential inclusion bias. Over-representation of certain groups

of individuals in the integrated dataset and differences in outcomes between groups could lead to biased results. While characterisation of the potential bias would not prevent error being introduced to subsequent analyses, understanding of the likely error could prevent false conclusions being drawn.

Once the bias of inclusion in the integrated dataset had been characterised, categories with small numbers or categories of missing information were considered for exclusion because results for these groups could not be meaningfully interpreted. Results for individuals without reported ethnicity or risk group would be difficult to interpret meaningfully but before exclusion it was expedient to determine whether this would introduce additional bias arising from the disproportionate exclusion of deaths. Inability to collect information on ethnicity or risk group was hypothesised to be due to incapacity on admission to HIV-related clinical care and subsequent death.

Differential CD4 cell count monitoring and LTFU were considered because these could directly bias subsequent results[247-254]. In addition, they may be associated with health and confounded with other effects such as adherence, psychosexual health, mental health and co-morbidities. Individuals with fewer CD4 cell counts, infrequent CD4 monitoring and more LTFU would be under-represented in the dataset. Studies have demonstrated how this can result in fewer observed events, loss of power and significant differences between predicted and observed outcomes[255-257]. Therefore, a statistical analysis of the factors associated with having only one CD4 cell count was undertaken along

with descriptive analysis of the number and frequency of CD4 cell counts for each individual and statistical analysis of temporary and permanent LTFU.

Finally, the proximity of CD4 cell count measurements to HIV diagnosis, AIDS diagnosis and death and factors associated with having a CD4 cell count within a defined period of these events were statistically characterised to help understand the effects of informative censoring. The delay from HIV diagnosis to first CD4 count could introduce bias in analyses of incidence due to left-censoring if follow-up time was considered to start at the first CD4 count and events before this time were excluded. The period between the last CD4 cell count and AIDS or death could introduce bias due to right-censoring of follow-up time and events. The results of these analyses were used in the analysis and interpretation of CD4 cell counts at the time of HIV diagnosis (chapter 6) and the incidence of AIDS and death (chapter 7).

These analyses provided interesting results about CD4 cell count monitoring and follow-up as well as a deeper understanding of the integrated dataset.

5.2 Aims

The following investigations were carried out to consider how the inclusion of records, the inclusion of events, the number and frequency of CD4 cell count monitoring and LTFU may bias the results of subsequent analyses.

a) To determine factors associated with the inclusion of records in the integrated dataset.

b) To determine whether individuals in the integrated dataset who had died were less likely to have ethnicity or risk group reported.

c) To describe the frequency and regularity of CD4 monitoring and to determine factors associated with:
    a. Having only one CD4 count;
    b. Having infrequent CD4 counts (a long median time between counts).

d) To determine factors that were independently associated with the incidence of permanent and temporary LTFU.

e) To describe the time between CD4 count measurements and HIV diagnosis, first AIDS diagnoses and deaths and to determine factors associated with having a CD4 cell count measured within a defined time from these events.

5.3 Methods

5.3.1 Factors associated with inclusion in the integrated dataset

A dataset was created containing single records with data from 87,521 individuals reported to HARS as diagnosed with HIV infection in England, Wales and Northern Ireland (E,W&NI) before 1st January 2008 and reported by the end of June 2008 (excluding 377 with no date of birth reported). These records were flagged if included in the integrated dataset and multivariable logistic regression was used to consider factors associated with inclusion. The factors considered were the period of HIV diagnosis, age group at HIV diagnosis, region of HIV diagnosis, risk group and ethnicity. Additionally, the Mantel-Haenszel estimate

of the odds ratio was used to determine whether death or a reported AIDS diagnosis were associated with inclusion in the integrated dataset. This analysis was stratified by the period of diagnosis for two reasons: 1) individuals diagnosed prior to 1995 must have survived to access care in 1995 to be included in SOPHID; 2) unreported deaths were likely to disproportionally reduce the proportion of records linked for individuals diagnosed prior to 1995.

5.3.2 Frequency of CD4 cell count monitoring

The integrated dataset consisted of separate records for each CD4 cell count measured between the time of HIV diagnosis and the end of 2007. Multivariable logistic regression was used to determine the factors associated with having only one CD4 count. The Mantel-Haenszel estimate of the odds ratio was used to determine whether death or a reported AIDS diagnosis were associated with having only one CD4 count. For individuals with more than one CD4 count, the Kaplan-Meier estimate was used to describe the overall time between consecutive CD4 counts. Descriptive analysis of the time between CD4 counts and descriptive analysis of the total number and frequency of CD4 counts was used to illustrate how CD4 count monitoring differed according to determining factors. A scatter plot overlaid with a linear regression plot of time between the first and last CD4 counts against the total number of CD4 counts was used to investigate the number of individuals with either a few CD4 counts over a long period of time or many CD4 counts over a short period of time.

5.3.3 Poisson regression analysis of loss to follow-up

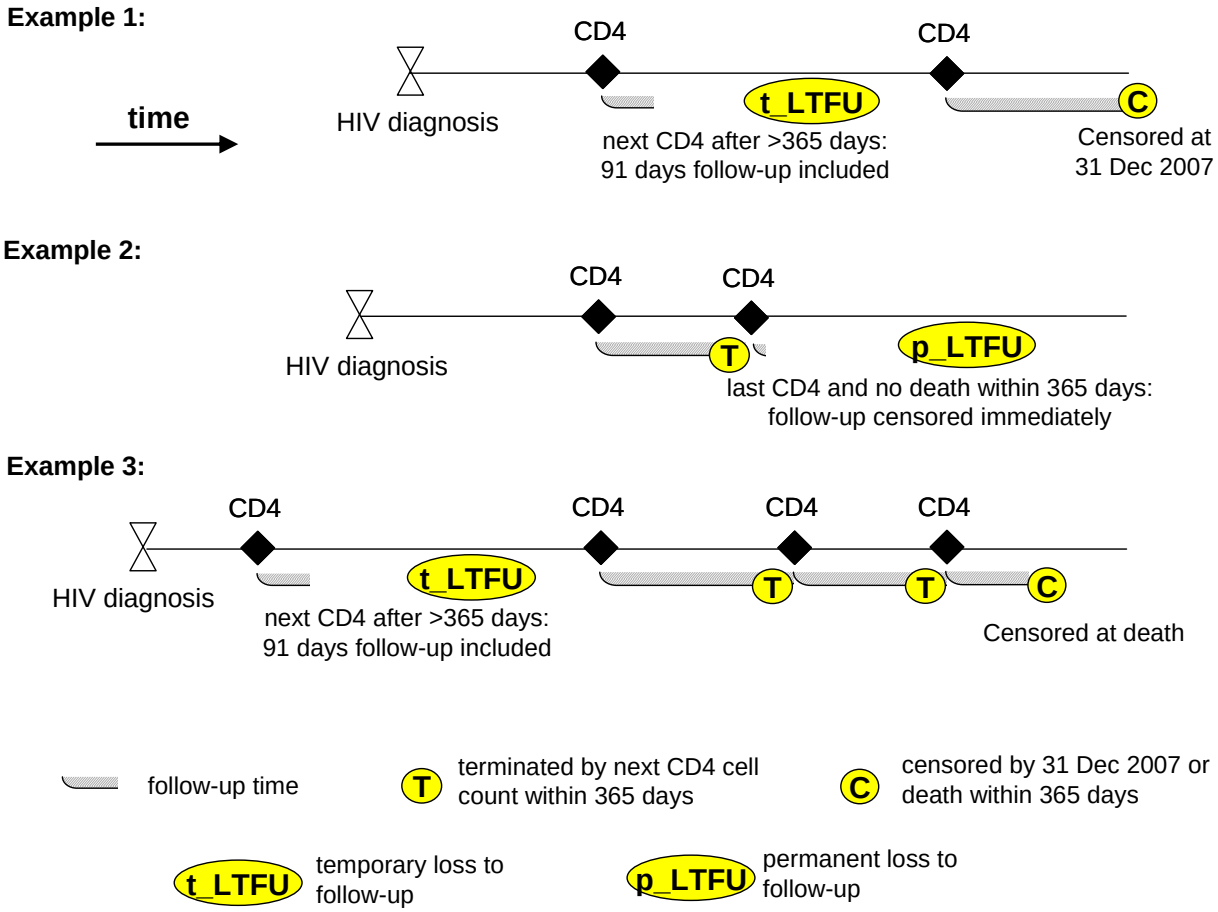
The period over which each individual was followed in the cohort was defined by separate CD4 counts in the integrated dataset (taking account of incomplete record linkage, LTFU, emigration and unreported deaths). Patient follow-up was split into distinct periods based on available CD4 cell counts (Figure 5.1).

In the analysis of temporary LTFU, each period of follow-up was assumed to start on the date of the CD4 cell count and end at the earliest of a) the next CD4 cell count, b) death, c) 31st December 2007, d) 91 days following the date of each CD4 count if the CD4 count was followed by another but not within 365 days. Follow-up was censored immediately after the last CD4 count if it was before the 1st January 2007 or if it was more than 365 days before the date of death. CD4 counts measured in 2007 were excluded from this analysis because of insufficient follow-up. Because individuals often have CD4 counts monitored quarterly[258;259], time at risk was expected to be continuous for most individuals although split into periods. However, follow-up censored because the subsequent CD4 count was measured after more than 365 days was defined as temporary LTFU (threshold informed by the Kaplan-Meier estimate of time between consecutive CD4 counts). If a patient was temporarily lost to follow-up 365 days after any particular CD4 count, they were allowed to re-enter the risk set if and when they had a further CD4 measurement.

Each period of follow-up, and losses to follow-up, were then allocated to a CD4 count category, age group and location (London or outside London) depending on the value of the CD4 count, the age at the time of the CD4 count and where

it was measured. This allowed these determining factors to change over time for each individual, but considered them to be constant during each period of follow-up. This ensured that person years of follow-up (PYFU) were included in the analysis with the appropriate determining factors. Finally, PYFU were allocated to calendar years in which they occurred. In the analysis of permanent LTFU, periods of follow-up were determined as above, but a last CD4 cell count before 1st January 2007 or no death reported within 365 days of the last CD4 cell count was considered as permanent LTFU.

Figure 5.1. Hypothetical calculation of periods of follow-up



Poisson regression was used for the analysis of incidence rates. This method assumes that rates are constant within time periods and estimates rate ratios between groups. The approach of allowing a patient's status to change permits the inclusion of variables that change over time. All PYFU are considered together where all determining variables are equal. This method is suited to analysis of trends as PYFU within the same calendar year are aggregated.

The main effects model included the following:

1. time-updated categorical variables: CD4 cell count categories, age groups, grouped calendar years
2. time-updated binary variables: previous AIDS diagnosis, region
3. fixed categorical variables: ethnicity; risk group (heterosexuals split by sex – numbers were not sufficient to split IDUs by sex).

5.3.4 Proximity of CD4 cell count measurements to events

Multivariable logistic regression was used to determine factors associated with having a CD4 cell count measured within a defined time from HIV diagnosis, to first AIDS diagnosis and to death. The defined time used to investigate these associations was informed by the Kaplan-Meier estimate of the time between a) each individual's date of HIV diagnosis and first CD4 cell count measurement; b) an individual's last CD4 cell count measurement and the date of death for individuals known to have died; c) an individual's last CD4 cell count measurement before their first AIDS diagnosis and the date of first AIDS diagnosis for individuals reported to have been diagnosed with AIDS.

5.4 Results

5.4.1 Factors of inclusion in the integrated dataset

The 50,953 individual records in the integrated dataset accounted for 58.2% of the total 87,521 HARS records (Table 5.1). Excluding those who were known to have died before 1995 who could not have been reported to SOPHID, this percentage increased to 67.3% of 77,212. The lower rate of inclusion in the integrated dataset in 2005-2007 compared with 2000-2004 may have been because a few laboratories were not able to report CD4 cell counts for 2007 (Section 2.3.2). Records from outside London were less likely to be included (adjusted odds ratio [AOR] 0.59; 95% CI 0.57, 0.61).

MSM were more likely to be included in the integrated dataset than other individuals, except recipients of blood/blood products (AOR 1.00; 95% CI 0.88, 1.14) and heterosexual women diagnosed antenatally (AOR 1.02; 0.93, 1.12), after accounting for other factors. Recipients of blood/blood products were poorly represented in the integrated dataset because most of these individuals were diagnosed before 1990 and less likely to survive to be included in SOPHID. There was also a higher proportion of younger individuals (AOR for those aged 15-25 years 1.41; 95% CI 1.33, 1.49 compared to those aged 30-34 years) and white individuals (AOR for black Africans 0.73; 95% CI 0.69, 0.76 in comparison) included in the integrated dataset.

Table 5.1. Number and percentage of HARS records in the integrated dataset

	Number of HARS patients included n (distribution, %)	All newly diagnosed patients n (distribution, %)	Percentage of newly diagnosed patients included %	Multivariable adjusted odds ratio (95% confidence interval)
Period of HIV diagnosis				
1979-1984	248 (0.5)	1,258 (1.6)	**20**	**0.10 (0.08, 0.12)**
1985-1989	2,504 (5)	11,243 (15)	**22**	**0.14 (0.13, 0.15)**
1990-1994	5,230 (10)	12,196 (16)	**43**	**0.40 (0.38, 0.42)**
1995-1999	8,851 (17)	13,004 (17)	68	0.97 (0.92, 1.02)
2000-2004	20,352 (40)	28,376 (37)	**72**	**1.26 (1.21, 1.32)**
2005-2007	13,768 (27)	21,444 (28)	64	-
Age group at HIV diagnosis				
15-25	6,382 (13)	10,800 (14)	**59**	**1.41 (1.33, 1.49)**
25-29	10,662 (21)	18,161 (24)	**59**	**1.14 (1.09, 1.20)**
30-34	11,889 (23)	20,233 (26)	59	-
35-39	9,457 (19)	16,027 (21)	59	0.98 (0.94, 1.03)
40-44	5,673 (11)	9,910 (13)	57	0.94 (0.89, 1.00)
45-49	3,158 (6)	5,624 (7)	**56**	**0.91 (0.85, 0.97)**
50-54	1,846 (4)	3,206 (4)	58	0.97 (0.89, 1.05)
>54	1,886 (4)	3,560 (5)	**53**	**0.77 (0.71, 0.83)**
Region of diagnosis				
Outside London	21,878 (43)	39,393 (51)	**56**	**0.59 (0.57, 0.61)**
London	29,075 (57)	48,128 (62)	60	-
Risk group				
MSM	22,984 (45)	40,186 (52)	57	-
Heterosexual men	9,221 (18)	15,046 (19)	**61**	**0.86 (0.81, 0.90)**
Heterosexual women	13,355 (26)	20,616 (27)	**65**	**0.93 (0.88, 0.98)**
Heterosexual women (diagnosed antenatally)	2,403 (5)	3,320 (4)	72	1.02 (0.93, 1.12)
IDU	1,621 (3)	3,533 (5)	**46**	**0.85 (0.78, 0.92)**
Recipients of blood products	487 (1)	1,715 (2)	28	1.00 (0.88, 1.14)
Children infected vertically	29 (0.1)	43 (0.1)	67	0.67 (0.35, 1.29)
Not reported	853 (2)	3,062 (4)	**28**	**0.51 (0.46, 0.56)**
Ethnicity				
White	25,870 (51)	41,693 (54)	62	-
Black African	18,594 (36)	27,936 (36)	**67**	**0.73 (0.69, 0.76)**
Black Caribbean	1,700 (3)	2,439 (3)	**70**	**0.81 (0.73, 0.89)**
Black Other	680 (1)	1,079 (1)	**63**	**0.73 (0.64, 0.84)**
Indian\Pakistani\Bangladeshi	722 (1)	1,154 (1)	**63**	**0.83 (0.73, 0.95)**
Other Asian	89 (0.2)	113 (0.1)	79	1.27 (0.80, 2.00)
Other/mixed	2,137 (4)	3,063 (4)	**70**	**0.88 (0.81, 0.96)**
Not reported	1,161 (2)	10,044 (13)	**12**	**0.10 (0.09, 0.11)**
Total	50,953 (100)	87,521 (100)	58	-

Stratification of included records by the period of HIV diagnosis indicated that the percentage of patients who were reported to have died decreased over time (Table 5.2). A lower proportion of records for individuals who died were included in the integrated dataset than for individuals who were not known to have died (26.8% [4,405/16,443] versus 65.5% [46,548/71,078]; p<0.01), irrespective of the period of HIV diagnosis. However, the univariable odds of being included in the integrated dataset increased after the earliest period of diagnosis. There was low inclusion of HARS patients who were diagnosed between 1985 and 1994 and who were not reported to have died.

Table 5.2. Percentage and number of HARS records that were included in the integrated dataset according to whether individuals had died or not.

Period of HIV Diagnosis	Percentage of HARS patients who had died % (n)	Percentage of HARS patients who had died that were included % (N)	Percentage of HARS patients who had <u>not</u> died that were included % (N)	Univariable odds ratio for each period (95% confidence interval)
1979-1984	76.6 (964)	11.8 (964)	45.6 (294)	**0.16 (0.12, 0.22)**
1985-1989	57.0 (6,408)	14.0 (6,408)	33.2 (4,835)	**0.33 (0.30, 0.36)**
1990-1994	39.9 (4,867)	29.4 (4,867)	51.8 (7,329)	**0.39 (0.36, 0.42)**
1995-1999	14.4 (1,871)	50.6 (1,871)	71.0 (11,133)	**0.42 (0.38, 0.46)**
2000-2004	5.7 (1,614)	46.2 (1,614)	73.3 (26,762)	**0.31 (0.28, 0.35)**
2005-2007	3.4 (719)	37.4 (719)	65.1 (20,725)	**0.32 (0.27, 0.37)**
Overall	18.8 (16,443)	26.8 (16,443)	65.5 (71,078)	**0.19 (0.19, 0.20)**

Categorisation of individuals who had died according to whether they had died within 31 days of their HIV diagnosis or not indicated that overall, 14.1% of individuals who died, did so shortly after HIV diagnosis (Table 5.3). These individuals accounted for a relatively small proportion of individuals who had

died, who were diagnosed before 1995, and who were not included in the integrated dataset. However, deaths shortly after HIV diagnosis accounted for an increasing percentage of all individuals who had died and who were not included in the integrated dataset (62.4% for individuals diagnosed between 2005 and 2007).

Overall, 9.8% (15.7% of those diagnosed after 1994) of patients who had died within 31 days of diagnosis were included in the integrated dataset compared to 29.0% (68.6% of those diagnosed after 1994) of patients who had died more than 31 days after diagnosis (Table 5.3). None of the individuals who were diagnosed before 1995 and who had died within 31 days of diagnosis were included in the integrated dataset because they could not have been linked to SOPHID records.

Table 5.3. Percentage and number of HARS patients who were included in the integrated dataset according to whether they had died within 31 days of HIV diagnosis or not.

Period of HIV diagnosis	Percentage of HARS patients who died and were not included who died within 31 days of diagnosis % (N)	Percentage of HARS patients who died within 31 days of diagnosis that were included % (N)	Percentage of HARS patients who died more than 31 days after diagnosis that were included % (N)	Univariable odds ratio for inclusion in the dataset by time from diagnosis (95% confidence interval)
1979-1984	3.8 (850)	0 (32)	12.2 (932)	-
1985-1989	5.2 (5,511)	0 (285)	14.6 (6,123)	-
1990-1994	11.4 (3,435)	0 (390)	32.0 (4,477)	-
1995-1999	31.4 (924)	9.4 (320)	59.1 (1,551)	**0.07 (0.05, 0.11)**
2000-2004	48.4 (868)	18.0 (512)	59.3 (1,102)	**0.15 (0.11, 0.20)**
2005-2007	62.4 (450)	18.3 (344)	54.9 (375)	**0.18 (0.13, 0.27)**
Overall	14.1 (12,038)	9.8 (1,883)	29.0 (14,560)	**0.27 (0.23, 0.31)**

A decreasing percentage of patients had ever had AIDS (Table 5.4). Patients with AIDS were less likely than those without AIDS to be included in the integrated dataset if they were diagnosed before 1995. In contrast, a reported AIDS diagnosis was not associated with inclusion in the integrated dataset for patients diagnosed after 1994.

Table 5.4. Percentage and number of HARS records that were included in the integrated dataset according to whether individuals had a reported AIDS diagnosis or not.

Period of HIV diagnosis	Percentage of HARS patients who had AIDS %	Percentage of HARS patients who had AIDS that were included % (N)	Percentage of HARS patients who had <u>not</u> AIDS that were included % (N)	Univariable odds ratio for each period (95% confidence interval)
1979-1984	75.5 (950)	14.4 (950)	36.0 (308)	**0.30 (0.22, 0.41)**
1985-1989	58.1 (6,528)	19.7 (6,528)	25.9 (4,715)	**0.70 (0.64, 0.77)**
1990-1994	46.8 (5,703)	38.6 (5,703)	46.6 (6,493)	**0.72 (0.67, 0.77)**
1995-1999	24.8 (3,220)	68.2 (3,220)	68.0 (9,784)	1.01 (0.93, 1.10)
2000-2004	12.7 (3,608)	71.2 (3,608)	71.8 (24,768)	0.97 (0.90, 1.05)
2005-2007	8.4 (1,798)	63.5 (1,798)	64.3 (19,646)	0.97 (0.87, 1.07)
Overall	24.9 (21,807)	43.7 (21,807)	63.0 (65,714)	**0.46 (0.44, 0.47)**

5.4.2 Further exclusions from the integrated dataset

The small number of records of children infected vertically (29 [0.1%]; 250 CD4 cell count measurements) were excluded from all further analyses. This group was unlikely to provide statistically significant comparisons with other groups and were potentially too different to be merged with other groups of individuals.

Records without a reported ethnicity and records without a reported risk group accounted for 0.9% (464) and 0.7% (342) of the individuals in the integrated

dataset. These relatively small numbers were excluded from further analyses because results could not be interpreted but it was unlikely to add much bias.

5.4.3 Differential CD4 cell count monitoring and loss to follow-up

There were 50,167 individuals with 875,237 CD4 cell counts in the final integrated dataset after exclusions; a median of 13 (IQR 5, 24) per patient. There were 3,442 (6.7%) individuals who had only one CD4 cell count and one individual had a maximum of 154 CD4 counts. The number of CD4 counts measured increased each year, from 20 in 1986 to 115,577 in 2007.

5.4.3.1 Individuals with only one CD4 cell count

A third (33.1% [1,138]) of the 3,442 individuals with only one CD4 cell count had it measured in 2007 and a further 9.0% (309) in 2006. Many of these individuals would be expected to have further CD4 counts reported. A quarter (26.0% [600]) of the individuals with single counts before 2007 were known to have died and 80.3% (482) of these were known to have died in the same or subsequent year (decreasing from 53.2% in 1995 to 8.7% in 2006).

A first CD4 cell count of 350 cells/mm^3 or more was independently associated with having only one CD4 cell count in the integrated dataset compared to counts of 200-249 cells/mm^3 (AOR at 350-499 cells/mm^3 1.53; 95% CI 1.29, 1.81) (Table 5.5): Younger age at the time of HIV diagnosis was also associated with having only one CD4 cell count (AOR at 15-24 years 1.41; 95% CI 1.23, 1.62 compared to age group 35-39 years). Individuals diagnosed between 1995 and 1998 or after 2004 were more likely to have only one CD4 cell count and

particularly those diagnosed in 2007. Individuals diagnosed before 1995 were likely to have subsequent measurements as they also had to have been reported to SOPHID since 1995. Individuals diagnosed outside of London, MSM and heterosexual women diagnosed antenatally, and individuals of white ethnicity were less likely to have only one CD4 cell count.

Table 5.5. Percentage of all individuals who had only one CD4 count and who were not known to have died in the same or subsequent year.

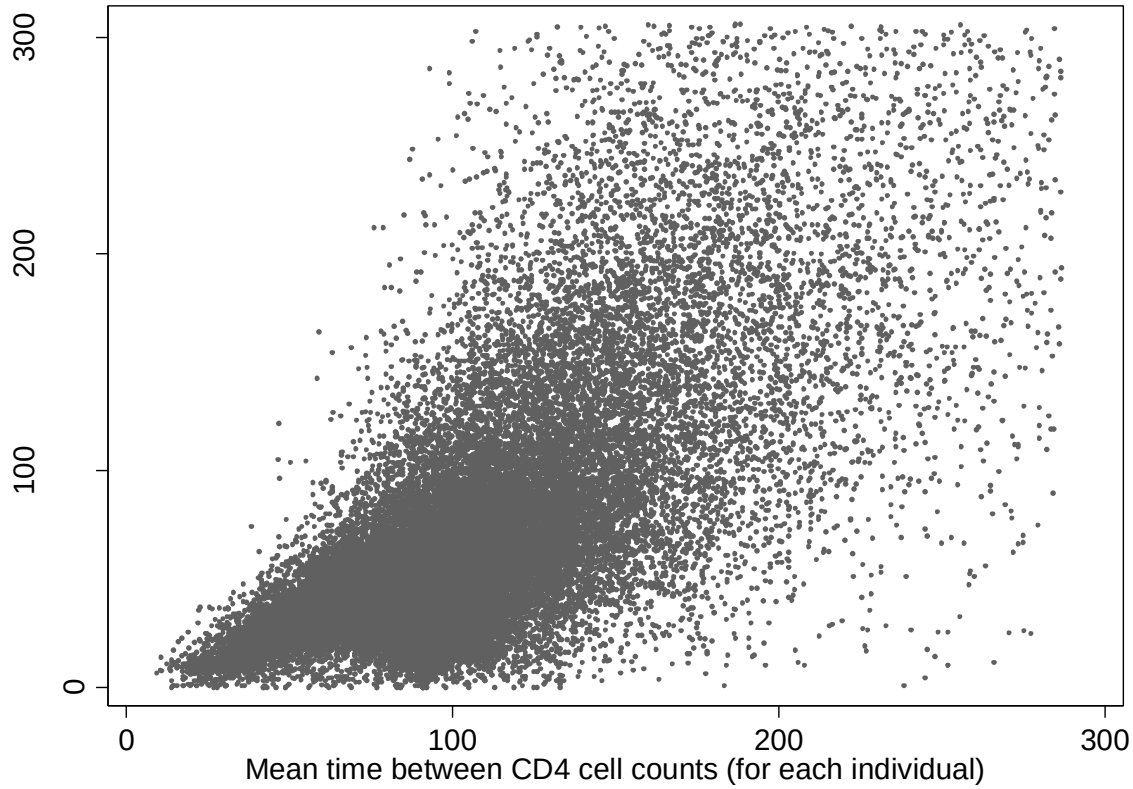
	Number of individuals with/without only one CD4 count	Percentage with only one CD4 count (not died in the same or subsequent year) (%)	Multivariable adjusted odds ratio (95% confidence interval)
Calendar year of HIV diagnosis			
<1991	94 / 3,338	**2.7**	**0.73 (0.56, 0.95)**
1991/1992	38 / 1,944	**1.9**	**0.51 (0.36, 0.74)**
1993/1994	59 / 2,352	**2.4**	**0.69 (0.51, 0.93)**
1995/1996	162 / 3,116	**4.9**	**1.43 (1.15, 1.78)**
1997/1998	168 / 3,487	**4.6**	**1.28 (1.03, 1.60)**
1999/2000	176 / 4,516	3.8	-
2001/2002	269 / 7,359	3.5	0.90 (0.74, 1.09)
2003/2004	421 / 9,507	4.2	1.08 (0.90, 1.30)
2005/2006	575 / 8,853	**6.1**	**1.62 (1.36, 1.93)**
2007	975 / 2,758	**26.1**	**9.21 (7.78, 10.92)**
Age group at HIV diagnosis			
15-24	467 / 6,146	**7.1**	**1.41 (1.23, 1.62)**
25-29	646 / 9,998	**6.1**	**1.18 (1.04, 1.34)**
30-34	607 / 11,069	**5.2**	**1.01 (0.89, 1.14)**
35-39	509 / 8,649	**5.6**	**-**
40-44	320 / 5,135	**5.9**	**1.00 (0.86, 1.16)**
45-49	176 / 2,881	**5.8**	**0.98 (0.81, 1.18)**
50-54	101 / 1,667	**5.7**	**1.08 (0.86, 1.36)**
>54	111 / 1,685	**6.2**	**1.14 (0.91, 1.43)**

First CD4 cell count category			
<50	247 / 5,449	4.3	0.86 (0.70, 1.05)
50-99	169 / 3,473	4.6	0.96 (0.77, 1.19)
100-149	157 / 3,283	4.6	0.92 (0.73, 1.15)
150-199	171 / 3,421	4.8	0.93 (0.75, 1.16)
200-249	195 / 3,742	5.0	-
250-299	203 / 3,697	5.2	1.02 (0.83, 1.26)
300-349	226 / 3,771	5.7	1.18 (0.96, 1.45)
350-499	710 / 9,182	**7.2**	**1.53 (1.29, 1.81)**
>499	859 / 11,212	**7.1**	**1.61 (1.36, 1.91)**
AIDS at HIV diagnosis			
No	2,607 / 39,518	6.2	-
Yes	330 / 7,712	4.1	0.92 (0.80, 1.06)
Region of diagnosis			
Outside London	1,247 / 20,430	5.8	**0.89 (0.82, 0.97)**
London	1,690 / 26,800	**5.9**	-
Risk group			
MSM	1,042 / 21,783	4.6	-
Heterosexual men	695 / 8,645	**7.4**	**1.56 (1.37, 1.78)**
Heterosexual women	909 / 12,584	**6.7**	**1.29 (1.13, 1.47)**
Heterosexual women (diagnosed antenatally)	130 / 2,271	5.4	0.81 (0.65, 1.01)
IDU	128 / 1,478	**8.0**	**2.36 (1.93, 2.88)**
Recipients of blood products	33 / 469	**6.6**	**2.17 (1.48, 3.18)**
Ethnicity			
White	1,239 / 24,886	4.7	-
Black-African	1,266 / 17,011	**6.9**	**1.42 (1.25, 1.61)**
Black Caribbean	123 / 1,528	**7.5**	**1.39 (1.13, 1.72)**
Black – other/unspecified	76 / 835	**8.3**	**1.48 (1.13, 1.92)**
Indian/Pakistani/Bangladeshi	43 / 609	6.6	1.18 (0.84, 1.65)
Other Asian	44 / 625	6.6	1.12 (0.81, 1.56)
Other/mixed	146 / 1,736	**7.8**	**1.47 (1.21, 1.78)**
Total	2,937 / 47,230	5.9	-

5.4.3.2 Frequency of CD4 cell count monitoring

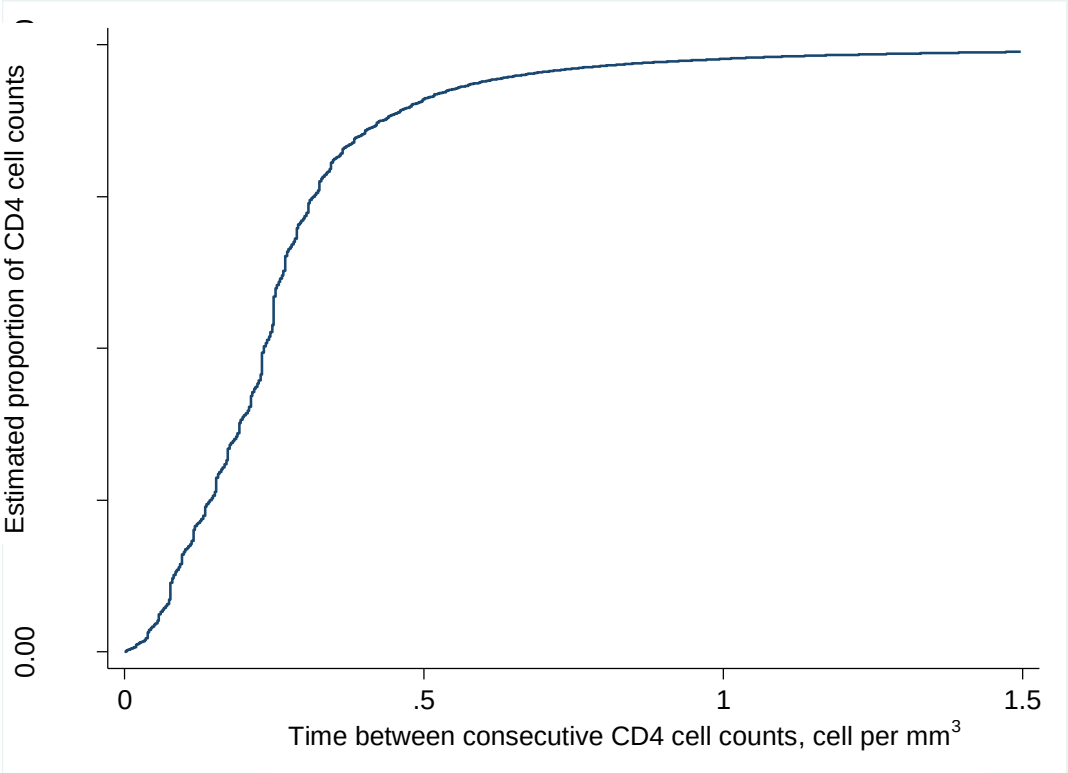
There were 46,725 individuals who had more than one CD4 cell count. The median for all individuals of the mean times between each individual's consecutive CD4 counts was 98 days (IQR 77, 131). The median standard deviation of time between CD4 counts for these individuals was 52 (IQR 34, 94) (Figure 5.2). Only few individuals had CD4 cell counts measured very regularly (low standard deviation in times between CD4 counts) and usually at one month or three month intervals

Figure 5.2. Plot of mean against standard deviation for time between CD4 counts (of each individual with more than two CD4 counts; excluding results with a mean or standard deviation above the 95th percentile).



The median time between any one CD4 count and the next was 85 days (IQR 53, 115). The Kaplan-Meier estimate showed that 41.5% (95% CI 41.4%, 41.6%), 9.0% (95% CI 8.9%, 9.0%) and 2.3% (95% CI 2.3%, 2.3%) of CD4 cell counts were followed by a subsequent measurement within 3 months, 6 months and a year (Figure 5.3).

Figure 5.3. Kaplan-Meier estimate of time between consecutive CD4 cell counts (for individuals with more than one CD4 cell count)



The median time between CD4 counts varied particularly by the CD4 count and also by whether the individual had a previous AIDS diagnosis (Table 5.6). The latter appeared to be a temporary effect as individuals appeared to have had shorter median times between consecutive counts (58 days [IQR 31, 92]) in the year of AIDS diagnosis than before or after (Table 5.7).

CD4 counts appeared to have been measured increasingly less frequently since the mid-1990s. Both the median follow-up between first and last CD4 cell counts and the overall number of CD4 cell counts fell over time since the first CD4 cell count due to censoring. However, there were relatively few CD4 cell counts measured for individuals whose first CD4 cell count was prior to 1995.

In descriptive analysis, risk group appeared to have relatively little association with the time between consecutive CD4 counts but substantial association with follow-up due to the timing of diagnosis of different groups. However, IDUs (91 days between consecutive CD4 counts [IQR 55, 133]) were monitored least frequently and MSM (84 days [IQR 50, 113]) most frequently.

Ethnicity had relatively little association with the frequency or length of follow-up. However, black Caribbean individuals (91 days between consecutive CD4 counts [IQR 57, 123]) were monitored slightly less frequently than individuals of other ethnicity (85 days [IQR 52, 115]). Individuals of white ethnicity had the longest follow-up (4.8 years [IQR 2.1, 9.2]) but also the most CD4 cell counts (17 [IQR 8, 32]). There was no substantial effect of age.

Table 5.6. Median times between consecutive CD4 counts and between first and last CD4 count (only individuals with more than one CD4 count).

	Median times between consecutive CD4 counts (IQR) /days	Median times between first and last CD4 counts (IQR) /years	Median overall number of CD4 counts (IQR)
Year of CD4 cell count	(of each CD4 count)	(of first CD4 count)	(of first CD4 count)
<1991	101 (70, 175)	16.8 (7.6, 17.5)	46 (25, 67)
1991/1992	84 (43, 124)	15.0 (5.4, 15.8)	41 (19, 60)
1993/1994	82 (42, 127)	12.8 (3.0, 13.7)	31 (10, 49)
1995/1996	70 (35, 124)	11.1 (6.1, 11.8)	34 (16, 46)
1997/1998	77 (43, 117)	9.4 (8.0, 10.2)	31 (18, 40)
1999/2000	82 (49, 111)	7.4 (6.5, 8.0)	24 (14, 31)
2001/2002	84 (51, 111)	5.3 (4.6, 6.0)	19 (11, 24)
2003/2004	88 (56, 113)	3.4 (2.8, 4.1)	13 (8, 17)
2005/2006	91 (61, 119)	1.6 (1.1, 2.2)	7 (5, 10)
2007	89 (56, 112)	0.4 (0.2, 0.6)	3 (2, 4)
Age group	(at each CD4 count)	(at first CD4 count)	(at first CD4 count)
15-24	87 (49, 125)	3.6 (1.5, 6.9)	12 (5, 23)
25-29	86 (49, 122)	4.1 (1.7, 7.9)	14 (6, 26)
30-34	85 (50, 119)	4.2 (1.9, 7.7)	15 (7, 27)
35-39	86 (53, 115)	4.0 (1.8, 7.0)	14 (7, 25)
40-44	86 (55, 113)	3.6 (1.5, 6.6)	14 (6, 25)
45-49	85 (54, 112)	3.4 (1.4, 6.6)	14 (6, 25)
50-54	85 (56, 111)	3.9 (1.7, 7.0)	15 (7, 27)
>54	85 (56, 108)	3.4 (1.3, 6.2)	13 (6, 25)
CD4 cell count category	(at each CD4 count)	(at first CD4 count)	(at first CD4 count)
<50	50 (28, 90)	3.5 (1.4, 6.4)	15 (7, 26)
50-99	56 (30, 91)	3.8 (1.7, 6.9)	15 (7, 27)
100-149	63 (35, 96)	3.8 (1.7, 6.8)	15 (7, 26)
150-199	67 (36, 98)	3.6 (1.6, 6.7)	14 (7, 25)
200-249	75 (42, 102)	3.9 (1.8, 7.2)	15 (7, 26)
250-299	83 (49, 106)	3.8 (1.6, 6.9)	14 (6, 25)
300-349	87 (56, 113)	4.0 (1.7, 7.5)	14 (7, 26)
350-499	91 (63, 119)	3.9 (1.7, 7.6)	13 (6, 26)
>499	96 (75, 127)	4.2 (1.8, 8.0)	12 (5, 24)
Previous AIDS diagnosis	(before each CD4 count)	(before first CD4 count)	(before first CD4 count)
No	89 (56, 119)	3.8 (1.7, 7.2)	13 (6, 25)
Yes	84 (49, 107)	4.3 (1.8, 7.5)	17 (8, 29)
Region of CD4 cell count	(each CD4 count)	(first CD4 count)	(first CD4 count)
Outside London	85 (56, 112)	3.3 (1.5, 5.9)	12 (6, 22)
London	86 (51, 119)	4.5 (1.9, 8.2)	15 (6, 28)

Risk group			
MSM	84 (50, 113)	4.8 (2.1, 9.4)	17 (8, 33)
Heterosexual men	89 (56, 116)	3.2 (1.3, 5.7)	11 (5, 20)
Heterosexual women	89 (56, 115)	3.3 (1.5, 5.7)	12 (6, 21)
Heterosexual women (diagnosed antenatally)	88 (49, 119)	2.8 (1.2, 4.6)	9 (5, 16)
IDU	91 (55, 133)	4.4 (1.7, 8.9)	12.5 (5, 25)
Recipients of blood products	85 (55, 114)	7.0 (2.7, 13.2)	22 (9, 44)
Ethnicity			
White	85 (51, 114)	4.8 (2.1, 9.2)	17 (8, 32)
Black-African	88 (56, 115)	3.1 (1.3, 5.2)	11 (5, 19)
Black Caribbean	91 (57, 123)	3.4 (1.4, 6.0)	11 (5, 20)
Black – other/unspecified	88 (52, 122)	3.6 (1.3, 6.8)	11 (5, 22)
Indian/Pakistani/Bangladeshi	89 (56, 117)	3.6 (1.5, 7.2)	13.5 (6, 26)
Other Asian	85 (56, 113)	3.1 (1.2, 5.9)	12 (6, 22)
Other/mixed	85 (52, 119)	3.6 (1.4, 7.4)	13 (5, 26)
Total	85 (53, 115)	3.9 (1.7, 7.3)	14 (6, 26)

Table 5.7. Median times between consecutive CD4 counts in the five years before and after a first AIDS diagnosis

Years before and after first AIDS diagnosis	Number of CD4 cell counts	Median (IQR) /days
-5	4,873	89 (51, 135)
-4	6,864	85 (49, 128)
-3	9,493	86 (49, 128)
-2	13,285	84 (48, 122)
-1	18,264	78 (42, 118)
0	**46,818**	**58 (31, 92)**
1	46,695	76 (42, 102)
2	37,669	84 (50, 107)
3	32,374	84 (55, 111)
4	27,952	84 (53, 112)
5	22,979	85 (56, 112)

For those individuals with a reported date of starting ART (startART), there was also a temporarily increased frequency of CD4 cell count monitoring (56 days [IQR 31, 91] between counts) in the year of startART (Table 5.8).

Table 5.8. Median times between consecutive CD4 counts in the five years before and after starting ART

Years before and after startART	Number of CD4 cell counts	Median (IQR) /days
-5	7,067	98 (63, 154)
-4	10,201	95 (63, 146)
-3	14,349	93 (63, 140)
-2	20,560	91 (61, 135)
-1	32,942	84 (49, 119)
0	**100,145**	**56 (31, 91)**
1	82,608	84 (55, 105)
2	65,454	90 (61, 112)
3	54,854	90 (61, 113)
4	45,196	90 (61, 114)
5	37,185	90 (61, 114)

There was also evidence that the first CD4 count measurement was followed relatively quickly by another CD4 count while subsequent CD4 counts were measured less frequently (Figure 5.4). Further analysis showed that this frequent monitoring of CD4 counts at the initiation of care was carried out irrespective of the level of first CD4 count (Figure 5.5). Individuals with very low first CD4 counts (CD4 <100 cells/mm^3) continued to be monitored frequently but gradually less so over time.

Figure 5.4. Median (IQR) times between consecutive CD4 counts according to their chronological order (for up to the first 30 CD4 counts of an individual)

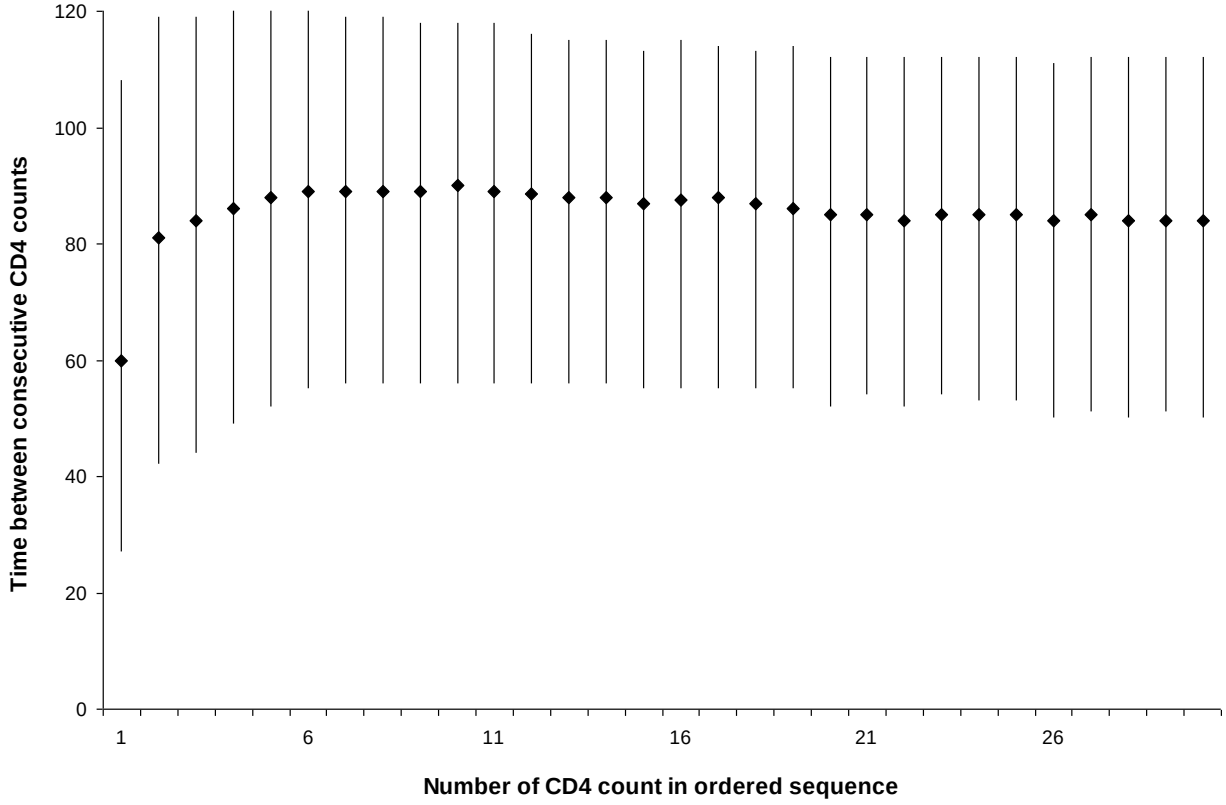
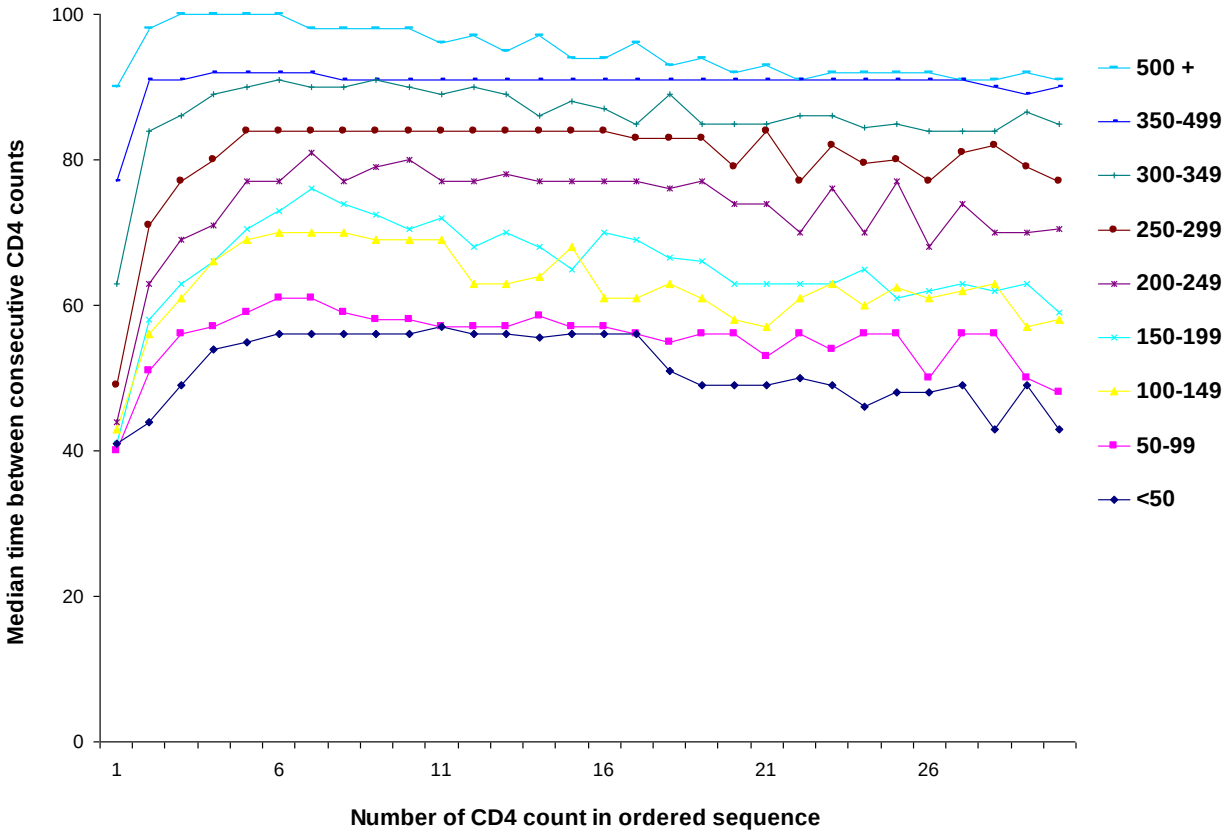
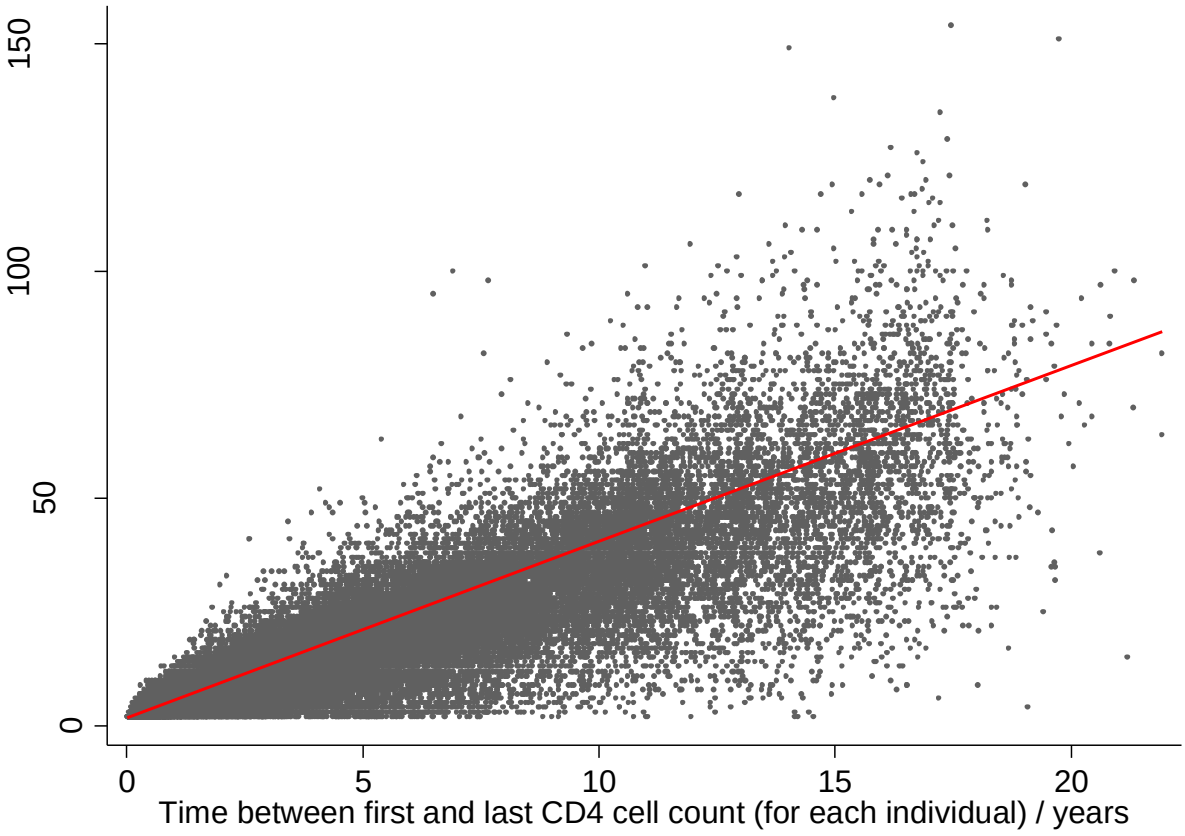


Figure 5.5. Median times between consecutive CD4 counts by CD4 cell count category (for up to the first 30 CD4 counts of an individual)



There were only a very small number of records in the integrated dataset for individuals with either a few CD4 cell count measurements over a long period of time or many CD4 cell count measurements over a short period of time (Figure 5.6).

Figure 5.6. Scatter plot and linear regression plot of time between the first and last CD4 cell counts against the total number of CD4 cell counts (of each individual with more than one CD4 cell count).



5.4.3.3 Temporary LTFU during care

Temporary LTFU, defined as subsequent CD4 cell counts measured after 365 days or more, occurred after 17,079 (2.1%) CD4 counts and for 12,246 (26.2%) individuals. The median CD4 cell count just prior to a period of temporary LTFU was 390 (IQR 240, 561) cells/mm^3. The rate of temporary LTFU decreased markedly over time (AOR in 1991/1992 2.07 and AOR in 2005/2006 0.47 compared to 2001/2002) (Table 5.9). Temporary LTFU also decreased with age

(AOR at 15-24 years 1.36 and AOR at >54 years 0.65 compared to 35-39 years). Patients who had previously been diagnosed with AIDS were less likely to be temporarily LTFU (AOR 0.68). Rates of temporary LTFU were higher at low CD4 cell counts (AOR at less than 50 cells/mm^3 1.30) and high CD4 cell counts (AOR at greater than 499 cells/mm^3 1.27) compared to CD4 cell counts of 200-249 cells/mm^3. MSM had a lower rate of temporary LTFU than heterosexuals (AOR for heterosexual men 1.40 and AOR for heterosexual women 1.09) and IDUs (AOR 1.91), but a higher rate than recipients of blood/blood products (AOR 0.78). Black individuals had a higher rate of temporary LTFU than white individuals after adjusting for other factors (AOR for black Africans 1.26 and AOR for black Caribbeans 1.49).

Table 5.9. Factors associated with the rate of temporary LTFU (CD4 count followed by more than 365 days before the next CD4 cell count).

	Number of CD4 counts with >365 days to next CD4 count / PYFU	Rate of temporary LTFU (>365 days to next CD4 cell count) / %	Multivariable rate ratio (95% confidence interval)
Year of CD4 cell count			
<1991	271 / 856	**31.6**	**2.57 (2.27, 2.93)**
1991/1992	665 / 3,335	**19.9**	**1.71 (1.57, 1.86)**
1993/1994	1,193 / 6,988	**17.1**	**1.53 (1.43, 1.64)**
1995/1996	1,725 / 10,186	**16.9**	**1.62 (1.52, 1.72)**
1997/1998	1,926 / 15,082	**12.8**	**1.28 (1.21, 1.35)**
1999/2000	2,452 / 21,342	**11.5**	**1.15 (1.09, 1.21)**
2001/2002	2,998 / 28,716	10.4	-
2003/2004	3,188 / 43,691	**7.3**	**0.67 (0.63, 0.70)**
2005/2006	2,661 / 55,568	**4.8**	**0.44 (0.41, 0.46)**

Age group at CD4 count			
15-24	1,317 / 7,392	**17.8**	**1.71 (1.60, 1.82)**
25-29	3,066 / 21,258	**14.4**	**1.38 (1.32, 1.45)**
30-34	4,289 / 38,104	**11.3**	**1.16 (1.11, 1.22)**
35-39	3,743 / 43,508	8.6	-
40-44	2,298 / 32,519	**7.1**	**0.91 (0.86, 0.96)**
45-49	1,162 / 19,411	**6.0**	**0.80 (0.74, 0.85)**
50-54	645 / 11,278	**5.7**	**0.78 (0.72, 0.85)**
>54	559 / 12,296	**4.5**	**0.66 (0.60, 0.72)**
CD4 cell count category			
<50	726 / 6,479	**11.2**	**1.26 (1.15, 1.38)**
50-99	662 / 6,556	**10.1**	**1.16 (1.06, 1.28)**
100-149	891 / 9,136	**9.8**	**1.15 (1.05, 1.25)**
150-199	990 / 11,546	8.6	1.04 (0.95, 1.13)
200-249	1,194 / 14,637	8.2	-
250-299	1,316 / 16,643	7.9	0.98 (0.91, 1.06)
300-349	1,555 / 17,759	**8.8**	**1.10 (1.02, 1.18)**
350-499	4,077 / 46,337	**8.8**	**1.12 (1.05, 1.20)**
>499	5,668 / 56,672	**10.0**	**1.32 (1.24, 1.41)**
Previous AIDS diagnosis			
No	13,635 / 125,028	10.9	-
Yes	3,444 / 60,739	**5.7**	**0.59 (0.56, 0.61)**
Region of CD4 cell count			
Outside London	5,758 / 65,824	8.7	0.99 (0.96, 1.02)
London	11,321 / 119,943	9.4	-
Risk group			
MSM	8,798 / 105,686	8.3	-
Heterosexual men	2,615 / 26,264	**10.0**	**1.40 (1.32, 1.47)**
Heterosexual women	3,672 / 39,529	**9.3**	**1.09 (1.03, 1.15)**
Heterosexual women (diagnosed antenatally)	668 / 5,013	**13.3**	**1.44 (1.32, 1.58)**
IDU	1,047 / 6,139	**17.1**	**1.91 (1.79, 2.04)**
Recipients of blood products	279 / 3,135	**8.9**	**0.78 (0.69, 0.88)**
Ethnicity			
White	10,382 / 119,006	8.7	-
Black-African	4,809 / 48,021	**10.0**	**1.26 (1.20, 1.33)**
Black Caribbean	583 / 4,801	**12.1**	**1.49 (1.36, 1.62)**
Black – other/unspecified	304 / 2,932	**10.4**	**1.17 (1.04, 1.31)**
Indian/Pakistani/Bangladeshi	202 / 2,252	9.0	1.10 (0.95, 1.27)
Other Asian	159 / 1,999	8.0	1.08 (0.92, 1.26)
Other/mixed	640 / 6,754	**9.5**	**1.09 (1.01, 1.18)**
Total	17,079 / 185,768	9.2	-

5.4.3.4 Permanent LTFU during care

There were 45,678 individuals who had 759,660 CD4 cell counts measured before 2007 and 10,365 (22.7%) of these individuals were permanently lost to follow-up. The median CD4 cell count just prior to permanent LTFU was 369 (IQR 220, 540) cells/mm^3. Multivariable analysis indicated that the rate of permanent LTFU increased significantly over time and was particularly high in 2005/6 (AOR 2.01 compared to 2001/2) (Table 5.10). This may be partly artificial due to insufficient follow-up from 2005/6 to distinguish permanent LTFU from temporary LTFU. There was a relatively high rate of permanent LTFU in 1995/6 (AOR 1.23). Permanent LTFU was relatively high for younger patients (AOR at ages 15-24 years 1.98 and AOR at ages 25-29 years 1.61 compared to age group 35-39 years). Individuals with CD4 cell counts less than 200 cells/mm^3 had higher rates of LTFU and these rates were particularly high at CD4 cell counts less than 50 cells/mm^3 (AOR 2.66). Individuals who had been previously diagnosed with AIDS were less likely to be lost to follow-up (AOR 0.57). Individuals with CD4 cell counts measured outside London were more likely to be lost to follow-up (AOR 1.41). IDUs were more than twice as likely as MSM to be lost to follow-up (AOR 2.35). Heterosexuals and non-white individuals were also significantly more likely than MSM and white individuals to be lost to follow-up.

Table 5.10. Factors associated with the rate of permanent LTFU (last CD4 cell count before the end of 2006 and not known to have died within 365 days)

	Number of CD4 counts with last CD4 count / PYFU	Rate of permanent LTFU (including >365 days to death) %	Multivariable rate ratio (95% confidence interval)
Year of CD4 cell count			
<1991	8 / 856	**0.9**	**0.20 (0.10, 0.41)**
1991/1992	82 / 3,335	**2.5**	**0.61 (0.49, 0.77)**
1993/1994	255 / 6,988	**3.6**	**0.85 (0.74, 0.98)**
1995/1996	537 / 10,186	**5.3**	**1.23 (1.11, 1.36)**
1997/1998	507 / 15,082	**3.4**	**0.89 (0.80, 0.99)**
1999/2000	779 / 21,342	3.7	0.99 (0.91, 1.09)
2001/2002	1,148 / 28,716	4.0	-
2003/2004	2,074 / 43,691	**4.7**	**1.08 (1.00, 1.16)**
2005/2006	4,975 / 55,568	**9.0**	**2.01 (1.88, 2.14)**
Age group at CD4 count			
15-24	853 / 7,392	**11.5**	**1.98 (1.82, 2.14)**
25-29	1,876 / 21,258	**8.8**	**1.61 (1.52, 1.72)**
30-34	2,341 / 38,104	**6.1**	**1.19 (1.12, 1.26)**
35-39	2,181 / 43,508	5.0	-
40-44	1,437 / 32,519	**4.4**	**0.91 (0.85, 0.97)**
45-49	733 / 19,411	**3.8**	**0.83 (0.77, 0.91)**
50-54	410 / 11,278	**3.6**	**0.85 (0.77, 0.95)**
>54	534 / 12,296	4.3	1.03 (0.94, 1.14)
CD4 cell count category			
<50	656 / 6,479	**10.1**	**2.66 (2.39, 2.96)**
50-99	449 / 6,556	**6.8**	**1.63 (1.45, 1.84)**
100-149	518 / 9,136	**5.7**	**1.23 (1.10, 1.37)**
150-199	622 / 11,546	5.4	1.10 (0.99, 1.22)
200-249	761 / 14,637	5.2	-
250-299	877 / 16,643	5.3	0.97 (0.88, 1.07)
300-349	944 / 17,759	5.3	0.96 (0.87, 1.05)
350-499	2,444 / 46,337	5.3	0.96 (0.89, 1.04)
>499	3,094 / 56,672	5.5	1.05 (0.97, 1.14)
Previous AIDS diagnosis			
No	8,106 / 125,028	6.5	-
Yes	2,259 / 60,739	**3.7**	**0.57 (0.54, 0.59)**

Region of CD4 cell count			
Outside London	4,780 / 65,824	**7.3**	**1.41 (1.35, 1.47)**
London	5,585 / 119,943	4.7	-
Risk group			
MSM	3,492 / 105,686	3.3	-
Heterosexual men	2,372 / 26,264	**9.0**	**1.66 (1.55, 1.77)**
Heterosexual women	3,250 / 39,529	**8.2**	**1.26 (1.18, 1.35)**
Heterosexual women (diagnosed antenatally)	640 / 5,013	**12.8**	**1.38 (1.25, 1.53)**
IDU	486 / 6,139	**7.9**	**2.35 (2.14, 2.59)**
Recipients of blood products	125 / 3,135	4.0	1.06 (0.88, 1.27)
Ethnicity			
White	4,284 / 119,006	3.6	-
Black-African	4,830 / 48,021	**10.1**	**1.90 (1.78, 2.03)**
Black Caribbean	349 / 4,801	**7.3**	**1.60 (1.43, 1.79)**
Black – other/unspecified	205 / 2,932	**7.0**	**1.60 (1.39, 1.85)**
Indian/Pakistani/Bangladeshi	147 / 2,252	**6.5**	**1.53 (1.30, 1.81)**
Other Asian	115 / 1,999	**5.8**	**1.26 (1.04, 1.52)**
Other/mixed	435 / 6,754	**6.4**	**1.69 (1.53, 1.87)**
Total	10,365 / 185,768	5.6	-

5.4.4 Proximity of CD4 cell count measurements to events

5.4.4.1 HIV diagnosis and time to first CD4 cell count

The median time to first CD4 cell count after HIV diagnosis was 15 (IQR 4, 69) days. This was substantially higher for individuals diagnosed before 1990 because few CD4 cell counts were reported before this time (Figure 5.7). Percentages of individuals diagnosed since 2000 who had not had a CD4 cell count within three months, six months and a year were 10.1%, 6.1% and 3.1% respectively (Table 5.11). A substantial proportion of individuals diagnosed since 1995 had not had a CD4 cell count measurement by one month. However, after two months there was only a gradual increase in the proportion of individuals with a CD4 cell count measurement (Figure 5.8).

Figure 5.7. Kaplan-Meier estimate of time from HIV diagnosis to first CD4 cell count measurement (first five years and first six months shown separately)

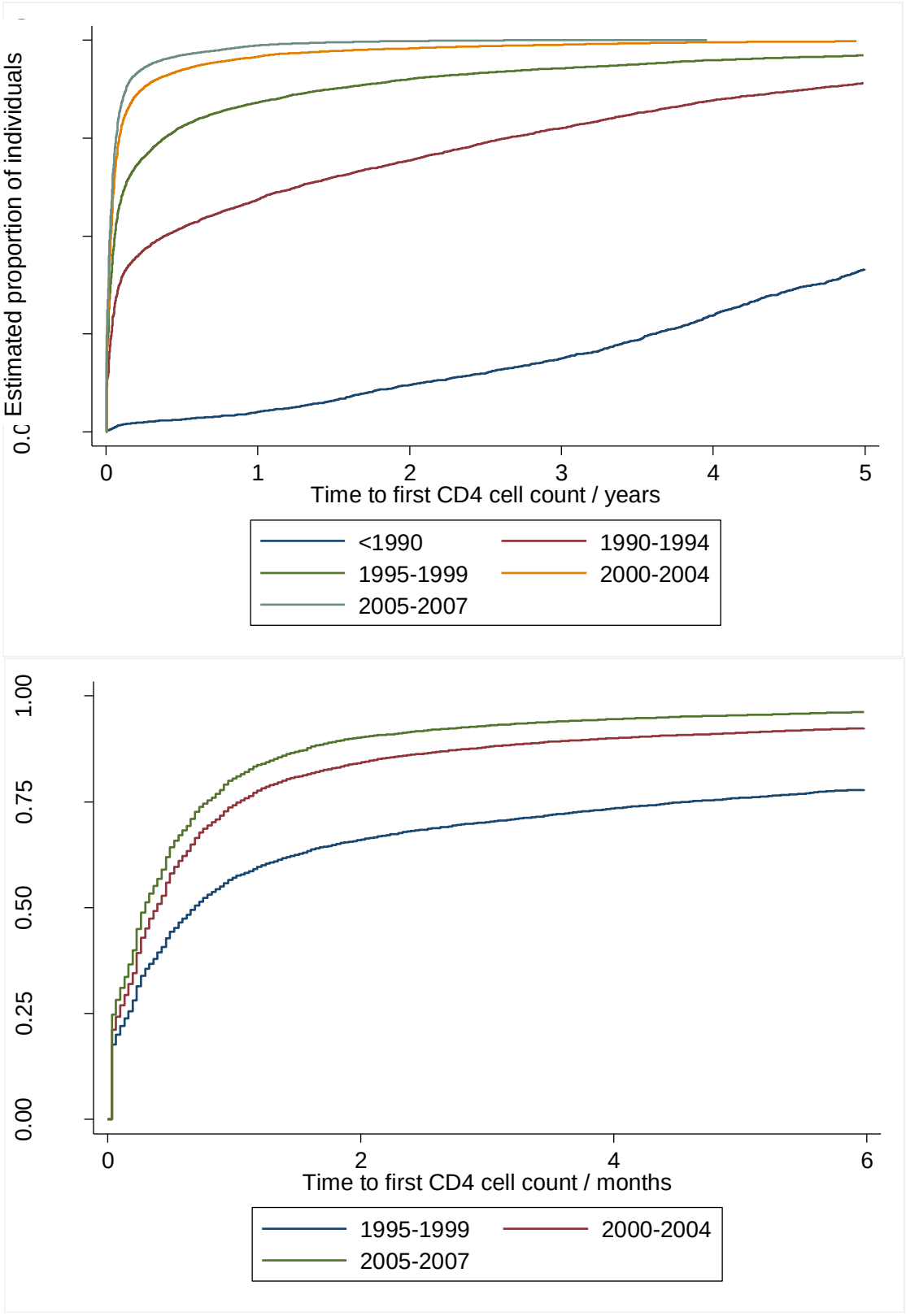


Table 5.11. Median time from HIV diagnosis to first CD4 count measurement and Kaplan-Meier estimate of proportion without a CD4 cell count after three, six and twelve months after HIV diagnosis, by period of HIV diagnosis

Period of HIV diagnosis	Median (IQR) /days	Kaplan-Meier estimate of proportion (95% CI) without a CD4 cell count within x months after HIV diagnosis / %		
		x = 3	x = 6	x = 12
<1990	2,078 (1,318, 2,955)	97.6 (96.9, 98.1)	96.9 (96.1, 97.5)	94.9 (94, 95.7)
1990-1994	140 (13, 963)	53.5 (52.1, 54.8)	47.9 (46.6, 49.3)	40.6 (39.3, 42)
1995-1999	21 (5, 140)	29.8 (28.9, 30.8)	22.1 (21.2, 22.9)	15.9 (15.2, 16.7)
2000-2004	12 (3, 32)	12.1 (11.6, 12.5)	7.6 (7.2, 8.0)	4.2 (4.0, 4.5)
2005-2007	9 (2, 24)	7.0 (6.6, 7.5)	3.8 (3.5, 4.2)	1.3 (1.2, 1.6)

Of individuals diagnosed since 1995, there were 85.7% (36,295/42,342) who had a CD4 cell count within 91 days of their HIV diagnosis (Table 5.12). The percentage of individuals with a CD4 cell count within 91 days of diagnosis increased substantially over time (from 32.4% in 1995/6 to 97.1% in 2007; AOR for 2007 12.69 compared to 1997/8). However, this was likely to be, in part, an artefact of censoring the data at the end of 2007. Age younger than 35 years at HIV diagnosis was independently associated with lower odds of having had a prompt CD4 cell count (AOR for ages 15-25 years 0.68 and AOR for ages 30-34 years 0.91 compared to ages 35-39 years). Individuals with AIDS at the time of HIV diagnosis were more likely to have had a prompt CD4 cell count (AOR 1.39) as were heterosexual women diagnosed antenatally when compared to MSM (AOR 1.83). Other heterosexuals (AOR for men 0.81 and AOR for women 0.87) and IDU (AOR 0.76) were less likely to have a prompt CD4 cell count than MSM. Individuals of black African (AOR 0.91) or black Caribbean (AOR 0.85) ethnicity were less likely to have prompt CD4 cell counts than white individuals.

Table 5.12. Factors associated with having a CD4 cell count measurement within 91 days of HIV diagnosis (individuals diagnosed since 1995)

	Number of individuals with/without a CD4 cell count within 91 days of HIV diagnosis	Percentage with a CD4 cell count within 91 days of HIV diagnosis (%)	Multivariable odds ratio (95% confidence interval)
Year of CD4 cell count			
1995/1996	2,047 / 1,231	**62.4**	**0.62 (0.56, 0.68)**
1997/1998	2,663 / 992	72.9	-
1999/2000	3,841 / 851	**81.9**	**1.69 (1.52, 1.87)**
2001/2002	6,522 / 1,106	**85.5**	**2.28 (2.06, 2.52)**
2003/2004	8,987 / 941	**90.5**	**3.73 (3.36, 4.13)**
2005/2006	8,611 / 817	**91.3**	**4.08 (3.68, 4.54)**
2007	3,624 / 109	**97.1**	**12.69 (10.34, 15.58)**
Age group at HIV diagnosis			
15-24	4,120 / 839	**83.1**	**0.68 (0.61, 0.75)**
25-29	7,076 / 1,408	**83.4**	**0.76 (0.69, 0.83)**
30-34	8,493 / 1,443	**85.5**	**0.91 (0.83, 0.99)**
35-39	7,103 / 1,039	87.2	-
40-44	4,246 / 595	87.7	0.96 (0.86, 1.08)
45-49	2,382 / 338	87.6	0.96 (0.83, 1.10)
50-54	1,386 / 198	87.5	1.03 (0.87, 1.22)
>54	1,489 / 187	88.8	1.14 (0.96, 1.35)
AIDS at HIV diagnosis			
No	30,591 / 5,238	85.4	
Yes	5,704 / 809	**87.6**	**1.39 (1.28, 1.51)**
Region of HIV diagnosis			
Outside London	16,508 / 2,499	86.9	0.96 (0.91, 1.02)
London	19,787 / 3,548	84.8	
Risk group			
MSM	14,818 / 2,549	85.3	
Heterosexual men	7,459 / 1,260	**85.5**	**0.81 (0.74, 0.90)**
Heterosexual women	10,764 / 1,823	**85.5**	**0.87 (0.79, 0.96)**
Heterosexual women (diagnosed antenatally)	2,219 / 166	**93.0**	**1.83 (1.53, 2.20)**
IDU	858 / 217	**79.8**	**0.76 (0.65, 0.90)**
Recipients of blood products	177 / 32	84.7	0.86 (0.58, 1.28)
Ethnicity			
White	16,781 / 3,018	84.8	
Black-African	15,041 / 2,368	**86.4**	**0.91 (0.83, 1.00)**
Black Caribbean	1,294 / 230	**84.9**	**0.85 (0.73, 1.00)**
Black – other/unspecified	716 / 98	88.0	1.15 (0.92, 1.45)
Indian/Pakistani/Bangladeshi	503 / 78	86.6	1.05 (0.81, 1.35)
Other Asian	566 / 62	90.1	1.24 (0.94, 1.63)
Other/mixed	1,394 / 193	**87.8**	**1.23 (1.04, 1.45)**
Total	36,295 / 6,047	85.7	-

There were some significant differences in the effects of the factors on the odds of having had a CD4 cell count measurement within 30, 91 and 182 days of HIV diagnosis (Table 5.13). It became clear that much of the association between diagnosis in 2007 and having had a CD4 cell count within the different time frames was artificial and due to censoring at the end of 2007. Heterosexual women diagnosed antenatally were less likely than MSM to have had a CD4 cell count within 30 days of HIV diagnosis (AOR 0.87) but more likely than MSM to have had a CD4 cell count within 91 or 182 days (AOR 1.99) of HIV diagnosis. Individuals older than 55 years were about a third more likely to have had a CD4 cell count measurement within 182 days than individuals aged 35-39 years (AOR 1.32) but this effect was not significant within a time frame of 30 or 91 days. Similarly, diagnosis outside London was associated with lower odds of having had a CD4 cell count within 182 days (AOR 0.89) but this association was not significant for the other time frames. Also, the association of black African and black Caribbean ethnicity were not associated with odds of having had a CD4 cell count within 182 days.

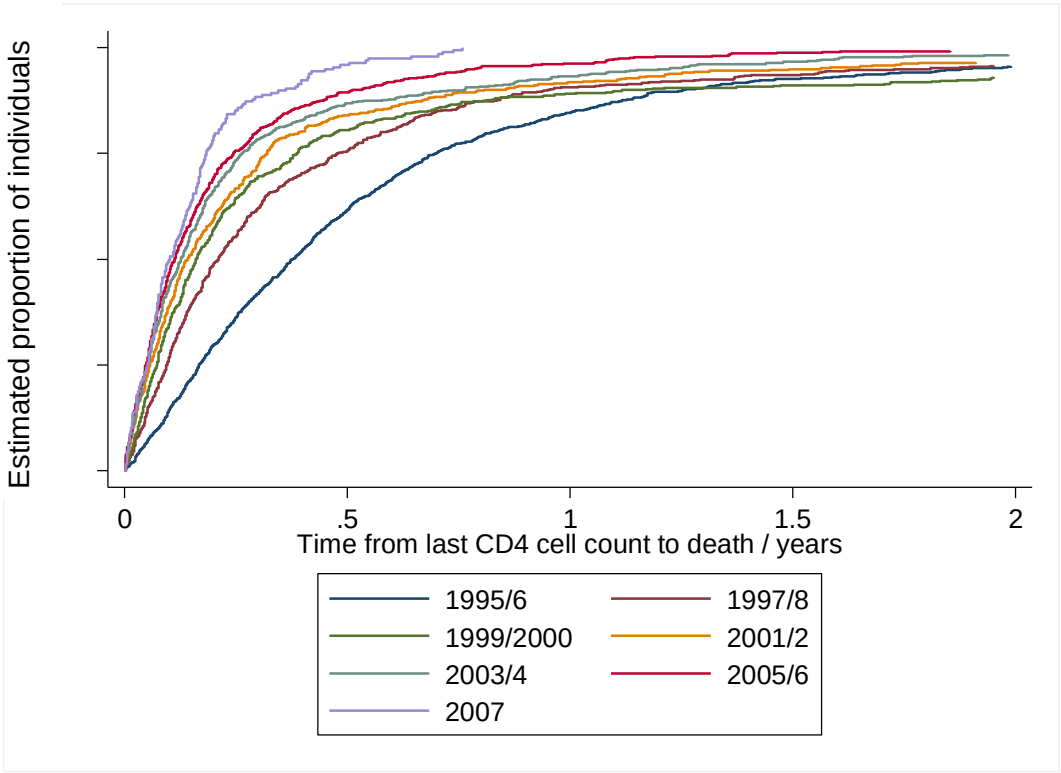
Table 5.13. Factors associated with having a CD4 cell count measurement within 30, 91 and 182 days of HIV diagnosis (individuals diagnosed since 1995)

	Multivariable odds ratio of having a CD4 cell count within x days of HIV diagnosis (95% confidence interval)		
	x = 30	**x = 91**	**x = 182**
Year of CD4 cell count			
1995/1996	**0.72 (0.66, 0.80)**	**0.62 (0.56, 0.68)**	**0.59 (0.53, 0.66)**
1997/1998	-	-	-
1999/2000	**1.47 (1.34, 1.61)**	**1.69 (1.52, 1.87)**	**1.70 (1.51, 1.92)**
2001/2002	**2.02 (1.86, 2.20)**	**2.28 (2.06, 2.52)**	**2.36 (2.11, 2.65)**
2003/2004	**2.35 (2.17, 2.56)**	**3.73 (3.36, 4.13)**	**4.58 (4.05, 5.19)**
2005/2006	**2.68 (2.47, 2.92)**	**4.08 (3.68, 4.54)**	**4.78 (4.22, 5.43)**
2007	**4.57 (4.07, 5.13)**	**12.69 (10.34, 15.58)**	**38.87 (25.78, 58.59)**
Age group at HIV diagnosis			
15-24	**0.76 (0.70, 0.82)**	**0.68 (0.61, 0.75)**	**0.68 (0.60, 0.77)**
25-29	**0.84 (0.78, 0.90)**	**0.76 (0.69, 0.83)**	**0.76 (0.68, 0.84)**
30-34	0.95 (0.89, 1.02)	**0.91 (0.83, 0.99)**	0.96 (0.87, 1.07)
35-39	-	-	-
40-44	1.00 (0.92, 1.08)	0.96 (0.86, 1.08)	1.04 (0.91, 1.19)
45-49	0.94 (0.85, 1.04)	0.96 (0.83, 1.10)	1.00 (0.85, 1.17)
50-54	1.03 (0.91, 1.17)	1.03 (0.87, 1.22)	1.07 (0.87, 1.31)
>54	1.02 (0.90, 1.15)	1.14 (0.96, 1.35)	**1.32 (1.06, 1.64)**
AIDS at HIV diagnosis			
No	-	-	-
Yes	**1.27 (1.19, 1.35)**	**1.39 (1.28, 1.51)**	**1.50 (1.36, 1.66)**
Region of HIV diagnosis			
Outside London	0.96 (0.92, 1.00)	0.96 (0.91, 1.02)	**0.89 (0.83, 0.95)**
London	-	-	-
Risk group			
MSM	-	-	-
Heterosexual men	**0.83 (0.77, 0.89)**	**0.81 (0.74, 0.90)**	**0.78 (0.70, 0.88)**
Heterosexual women	**0.86 (0.80, 0.93)**	**0.87 (0.79, 0.96)**	**0.87 (0.78, 0.98)**
Heterosexual women (diagnosed antenatally)	**0.87 (0.78, 0.98)**	**1.83 (1.53, 2.20)**	**1.99 (1.57, 2.51)**
IDU	**0.67 (0.59, 0.77)**	**0.76 (0.65, 0.90)**	**0.73 (0.61, 0.88)**
Recipients of blood products	0.85 (0.62, 1.15)	0.86 (0.58, 1.28)	0.92 (0.57, 1.50)
Ethnicity			
White	-	-	-
Black-African	**0.86 (0.80, 0.92)**	**0.91 (0.83, 1.00)**	0.91 (0.82, 1.02)
Black Caribbean	**0.77 (0.68, 0.87)**	**0.85 (0.73, 1.00)**	0.94 (0.78, 1.14)
Black – other/unspecified	0.87 (0.74, 1.02)	1.15 (0.92, 1.45)	1.07 (0.82, 1.40)
Indian/Pakistani/Bangladeshi	1.00 (0.82, 1.22)	1.05 (0.81, 1.35)	1.08 (0.80, 1.47)
Other Asian	1.21 (0.99, 1.48)	1.24 (0.94, 1.63)	1.06 (0.77, 1.46)
Other/mixed	1.12 (0.99, 1.27)	**1.23 (1.04, 1.45)**	**1.25 (1.02, 1.52)**

5.4.4.2 Death and time from last CD4 cell count

There were 4,564 individuals with a reported date of death in the integrated dataset. All of these deaths occurred since 1995 as all individuals in the integrated dataset had been reported to SOPHID. Therefore, individuals whose last CD4 cell count before death was prior to 1995 were excluded from the remainder of this analysis. For the remaining 4,179 deaths, the median time between the last CD4 cell count measurement and the date of death was 68 days (IQR 28, 164). This time between last CD4 cell count and death decreased with increasing period of last CD4 cell count (Figure 5.9). This was not simply due to decreasing follow-up to the end of 2007 because the Kaplan-Meier estimates changed little when only deaths within two years of the last CD4 cell count measurement were considered. Overall estimates of the percentages of individuals with more than three months, six months and a year between their last CD4 cell count and death were 42.2%, 23.3% and 10.1% respectively (excluding individuals with their last CD4 cell count in 2007 because they could not have had more than 365 days to their date of death).

Figure 5.9. Kaplan-Meier estimate of time from last CD4 count measurement to death (individuals with a last CD4 count measurement in 1995 or later)



Of individuals with their last measurement between 1995 and 2006, 90.0% (3,582/3,982) had a CD4 cell count within 365 days of death (Table 5.14). The median of the last CD4 cell count within 365 days of death was 55 (IQR 13, 190) cells/mm^3 compared to 100 (IQR 21, 300) cells/mm^3 where the last measurement was more than 365 days before death. In multivariable analysis, individuals with a CD4 cell count less than 50 cells/mm^3 (AOR 1.74) and individuals with a previous AIDS diagnosis (AOR 2.39) were more likely to have had a CD4 cell count within 365 days of death. Individuals aged older than 54 years also had increased odds of having had a CD4 cell count within 365 days of death (AOR 2.04). IDU (AOR 0.69) and individuals aged 25-29 years (AOR

0.64) were less likely to have had a CD4 cell count within 365 days of death. Individuals with their last CD4 cell count in 1995 or 1996 (AOR 0.34) were substantially less likely to have had a CD4 cell count within 365 days of death than those who had their last CD4 cell count measurement in 2001 or 2002, who in turn, were substantially less likely to have had a CD4 cell count within 365 days of death than those who had their last CD4 cell count measurement in 2005 or 2006 (AOR 2.75). IDU were less likely to have a CD4 cell count within 365 days of death than MSM (AOR 0.69). There was no evidence of an association between ethnicity and having had a CD4 cell count within 365 days of death.

Table 5.14. Factors associated with having a CD4 count measurement within 365 days of death (individuals' last measurement between 1995 and 2006)

	Number of individuals with/without a CD4 cell count within 365 days of death	Percentage with a CD4 cell count within 365 days of death (%)	Multivariable odds ratio (95% confidence interval)
Year of last CD4 cell count			
1995/1996	1,025 / 186	**84.6**	**0.34 (0.23, 0.51)**
1997/1998	522 / 54	90.6	0.78 (0.50, 1.23)
1999/2000	491 / 60	89.1	0.73 (0.47, 1.13)
2001/2002	437 / 39	91.8	-
2003/2004	537 / 39	93.2	1.40 (0.87, 2.26)
2005/2006	570 / 22	**96.3**	**2.75 (1.58, 4.77)**

Age group			
15-24	53 / 8	86.9	1.24 (0.54, 2.86)
25-29	274 / 61	**81.8**	**0.64 (0.44, 0.94)**
30-34	630 / 86	88.0	0.96 (0.69, 1.34)
35-39	736 / 87	89.4	-
40-44	641 / 59	91.6	1.21 (0.85, 1.74)
45-49	513 / 56	90.2	0.98 (0.68, 1.42)
50-54	299 / 20	93.7	1.54 (0.92, 2.60)
>54	436 / 23	**95.0**	**2.04 (1.24, 3.35)**
CD4 cell count category			
<50	1,679 / 139	**92.4**	**1.74 (1.02, 2.96)**
50-99	470 / 57	89.2	1.02 (0.58, 1.80)
100-149	336 / 31	91.6	1.25 (0.68, 2.33)
150-199	223 / 30	88.1	0.89 (0.47, 1.68)
200-249	166 / 20	89.2	-
250-299	143 / 22	86.7	0.73 (0.37, 1.43)
300-349	112 / 18	86.2	0.69 (0.34, 1.41)
350-499	224 / 40	84.8	0.58 (0.32, 1.06)
>499	229 / 43	84.2	0.61 (0.34, 1.11)
Previous AIDS diagnosis			
No	927 / 194	82.7	-
Yes	2,655 / 206	**92.8**	**2.39 (1.88, 3.04)**
Region of CD4 cell count			
Outside London	1,466 / 162	90.0	0.88 (0.70, 1.11)
London	2,116 / 238	89.9	-
Risk group			
MSM	2,076 / 222	90.3	-
Heterosexual men	577 / 57	91.0	0.82 (0.55, 1.23)
Heterosexual women	514 / 54	90.5	0.95 (0.61, 1.46)
Heterosexual women (diagnosed antenatally)	20 / 4	83.3	0.42 (0.13, 1.41)
IDU	302 / 54	**84.8**	**0.69 (0.49, 0.98)**
Recipients of blood products	93 / 9	91.2	1.16 (0.56, 2.41)
Ethnicity			
White	2,565 / 294	89.7	-
Black-African	704 / 76	90.3	0.86 (0.57, 1.32)
Black Caribbean	75 / 11	87.2	0.64 (0.31, 1.31)
Black – other/unspecified	40 / 3	93.0	1.54 (0.45, 5.29)
Indian/Pakistani/Bangladeshi	53 / 6	89.8	0.82 (0.33, 2.03)
Other Asian	28 / 2	93.3	1.01 (0.22, 4.54)
Other/mixed	117 / 8	93.6	1.47 (0.69, 3.12)
Total	3,582 / 400	90.0	-

5.4.4.3 First AIDS diagnosis and time from previous CD4 cell count

There were 16,616 individuals with a reported AIDS diagnosis in the integrated dataset. A substantial proportion of these diagnoses occurred at the time of HIV diagnosis and before a CD4 cell count had been measured such that there were only 10,266 (61.8%) AIDS diagnoses with a prior CD4 cell count measurement. Only 18.0% (873/1,797) of the AIDS diagnoses before 1995 had an earlier CD4 cell count measurement, compared to 63.1% (9,393/14,819) since 1995, and therefore individuals whose last CD4 cell count before AIDS was prior to 1995 were excluded from the remainder of this analysis. There were 9,087 AIDS diagnoses with a prior CD4 cell count measurement, which was in 1995 or later. Whereas 89.1% (7,236/8,118) of individuals with AIDS diagnoses more than three months after HIV diagnosis had a previous CD4 cell count measurement, this was only the case for 29.7% (1,851/6,222) of individuals with AIDS diagnoses within three months of their HIV diagnosis (Figure 5.10).

The median time between the last CD4 cell count measurement before AIDS and the date of AIDS diagnosis was 21 days (IQR 3, 66). This interval was shorter for individuals with AIDS within three months of their HIV diagnosis (median 1 day [IQR 0, 12]) than for individuals with AIDS more than three months after HIV diagnosis (median 33 days [IQR 9, 87]). For individuals with AIDS more than three months after HIV diagnosis, the time between the CD4 cell count and AIDS diagnosis decreased with increasing year of CD4 cell count (Figure 5.11), as for time to death, and this was not simply due to decreasing follow-up to the end of 2007. Overall percentages of individuals with more than

three months, six months and a year between their CD4 cell count and AIDS diagnosis were 23.5%, 11.8% and 5.9% respectively (excluding individuals with their CD4 cell count measurement in 2007 because they could not have had more than 365 days to their date of AIDS diagnosis).

Figure 5.10. Numbers of AIDS diagnoses categorised by time from HIV diagnosis and the previous CD4 cell count measurement

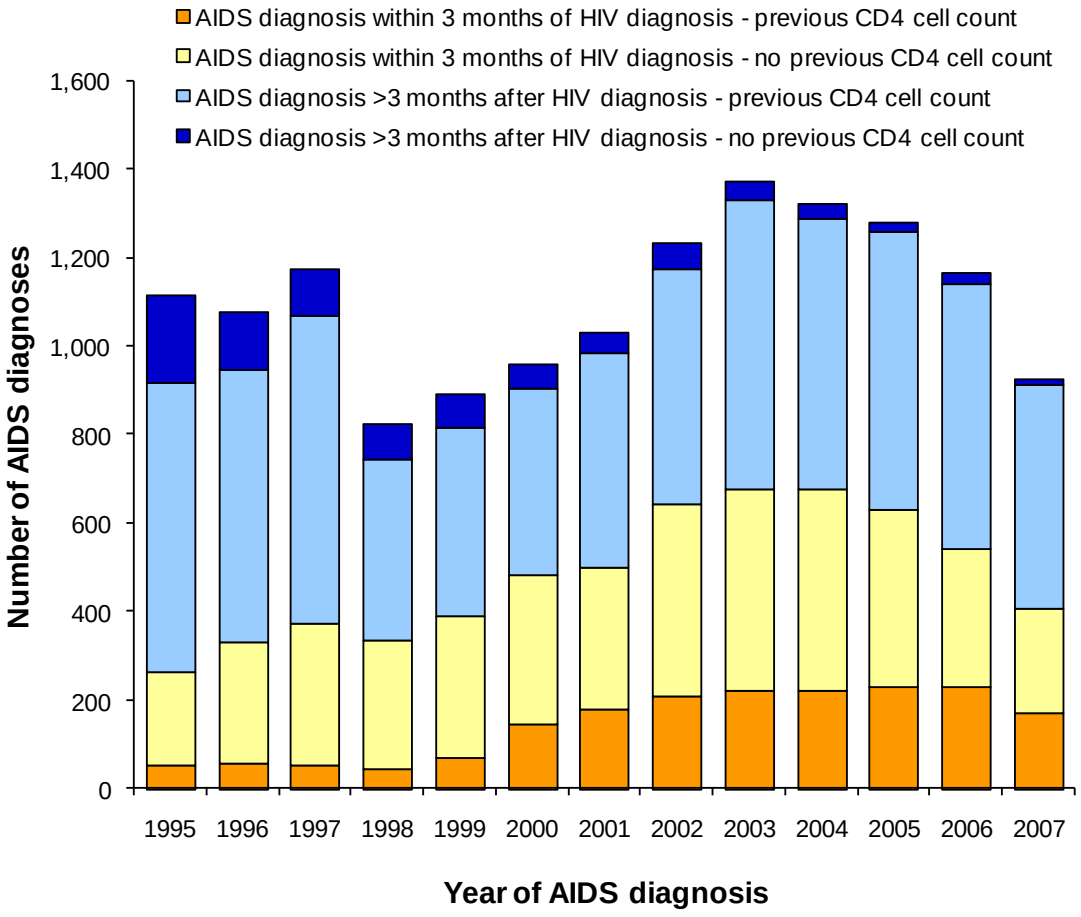
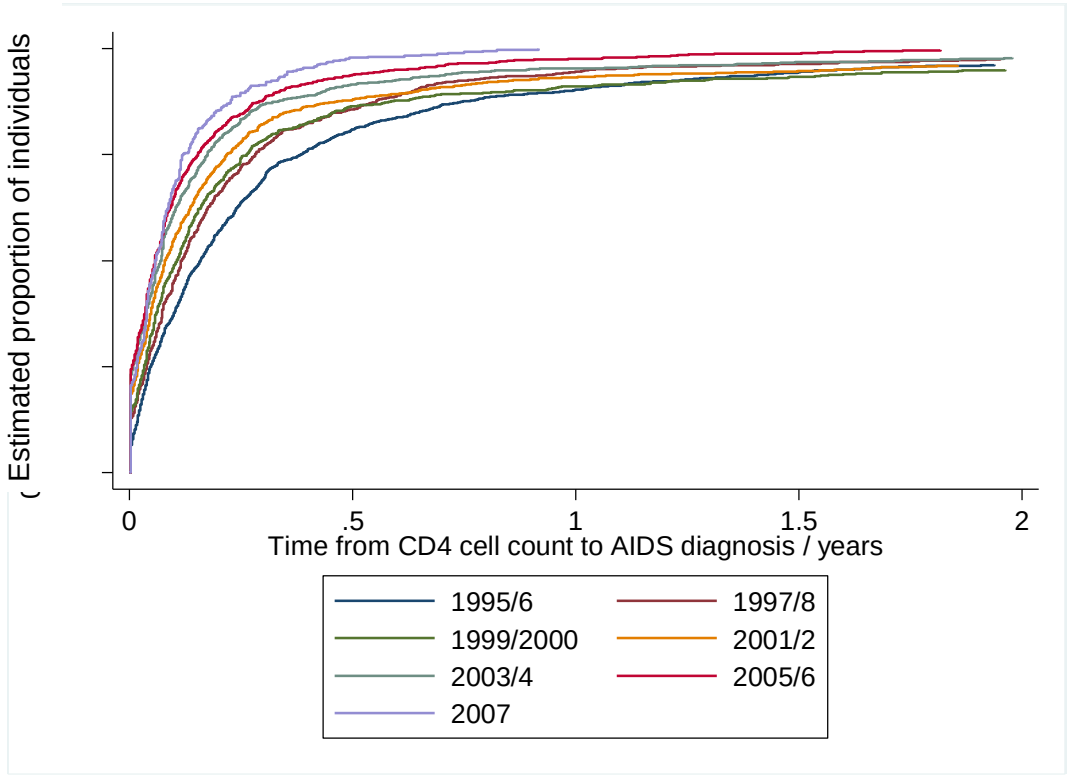


Figure 5.11. Kaplan-Meier estimate of time from CD4 count measurement to AIDS diagnosis (individuals with a last CD4 count in 1995 or later)



For individuals with their last CD4 cell count measurement between 1995 and 2006 and without AIDS at HIV diagnosis, there were a number of factors associated with the odds of having a CD4 cell count measurement within 365 days prior to a first AIDS diagnosis (Table 5.15). The odds of having a CD4 cell count measurement within 365 days prior to a first AIDS diagnosis was particularly high for individuals whose last CD4 cell count before AIDS was measured since 2003 when compared to 2001/2002 (AOR for 2003/4 1.63 and AOR for 2005/6 3.31), whereas these odds were low if the last CD4 cell count before AIDS was measured in 1995/6 (AOR 0.52) or 1999/2000 (AOR 0.70). There appeared to be an increasing likelihood of having a CD4 cell count

measurement within 365 days prior to a first AIDS diagnosis with increasing age (AOR for ages 15-24 years 0.56 and AOR for ages older than 54 years 2.38 compared to ages 35-39 years). Multivariable analysis indicated that the odds of having a CD4 cell count measurement within 365 days prior to a first AIDS diagnosis decreased with increasing CD4 cell count (AOR for CD4 cell count less than 50 cells/mm^3 2.01 and AOR for CD4 cell count greater than 499 cells/mm^3 0.49 compared to CD4 category 200-249 cells/mm^3). Individuals whose last CD4 cell count before AIDS was measured outside London were significantly less likely to have a CD4 cell count measurement within 365 days prior to a first AIDS diagnosis (AOR 0.60). There was no significant effect of ethnicity and little effect of risk group, although IDU were significantly less likely to have a CD4 cell count within 365 days prior to a first AIDS diagnosis (AOR 0.61).

Table 5.15. Factors associated with having a CD4 cell count measurement within 365 days of a first AIDS diagnosis (individuals with their last measurement between 1995 and 2006 and without AIDS at HIV diagnosis)

	Number of individuals with/without a CD4 cell count within 365 days of AIDS	Percentage with a CD4 cell count within 365 days of AIDS (%)	Multivariable odds ratio (95% confidence interval)
Year of last CD4 cell count			
1995/1996	1,314 / 141	**90.3**	**0.52 (0.37, 0.71)**
1997/1998	985 / 57	94.5	1.14 (0.78, 1.66)
1999/2000	795 / 77	**91.2**	**0.70 (0.49, 0.99)**
2001/2002	937 / 68	93.2	-
2003/2004	1,202 / 58	**95.4**	**1.63 (1.13, 2.35)**
2005/2006	1,151 / 28	**97.6**	**3.31 (2.10, 5.20)**

Age group			
15-24	230 / 29	**88.8**	**0.56 (0.35, 0.89)**
25-29	840 / 85	**90.8**	**0.70 (0.51, 0.95)**
30-34	1,460 / 117	92.6	0.84 (0.63, 1.13)
35-39	1,494 / 94	94.1	-
40-44	1,063 / 54	95.2	1.26 (0.89, 1.78)
45-49	605 / 24	96.2	1.59 (1.00, 2.54)
50-54	331 / 16	95.4	1.43 (0.82, 2.49)
>54	361 / 10	**97.3**	**2.38 (1.21, 4.66)**
CD4 cell count category			
<50	1,063 / 45	**95.9**	**2.01 (1.29, 3.12)**
50-99	760 / 32	**96.0**	**1.90 (1.17, 3.06)**
100-149	693 / 28	**96.1**	**1.85 (1.12, 3.04)**
150-199	643 / 39	94.3	1.13 (0.71, 1.78)
200-249	633 / 42	93.8	-
250-299	526 / 39	93.1	0.83 (0.53, 1.32)
300-349	450 / 38	92.2	0.74 (0.46, 1.18)
350-499	882 / 84	91.3	**0.63 (0.42, 0.93)**
>499	734 / 82	90.0	**0.49 (0.33, 0.72)**
Region of CD4 cell count			
Outside London	2,183 / 187	**92.1**	**0.60 (0.49, 0.74)**
London	4,201 / 242	94.6	-
Risk group			
MSM	3,325 / 219	93.8	-
Heterosexual men	1,068 / 59	94.8	0.87 (0.59, 1.28)
Heterosexual women	1,449 / 99	93.6	0.83 (0.57, 1.20)
Heterosexual women (diagnosed antenatally)	143 / 10	93.5	0.90 (0.43, 1.89)
IDU	307 / 36	**89.5**	**0.61 (0.41, 0.89)**
Recipients of blood products	92 / 6	93.9	1.08 (0.46, 2.54)
Ethnicity			
White	3,741 / 268	93.3	-
Black-African	2,001 / 122	94.3	0.94 (0.65, 1.37)
Black Caribbean	172 / 10	94.5	0.87 (0.44, 1.73)
Black – other/unspecified	98 / 7	93.3	0.84 (0.37, 1.90)
Indian/Pakistani/Bangladeshi	90 / 7	92.8	0.68 (0.30, 1.55)
Other Asian	66 / 5	93.0	0.71 (0.28, 1.85)
Other/mixed	216 / 10	95.6	1.33 (0.68, 2.59)
Total	6,384 / 429	93.7	-

5.5 Discussion

There was significant potential for bias due to differential inclusion, CD4 cell count monitoring, temporary and permanent LTFU, which may in turn bias the subsequent analyses.

5.5.1 Potential inclusion bias

A key determinant of bias, created by the formation of the integrated dataset, was the disproportionate exclusion of records for individuals diagnosed before 1995 that were not linked to SOPHID reports. This was primarily due to deaths before 1995, when there was a high mortality rate, including deaths that were not reported. Individuals included had to have survived long enough and remained in E,W&NI long enough to be reported to SOPHID[247]. This bias was likely to be the main cause of the following observations: a) permanent LFTU was low before 1993; b) individuals diagnosed before 1995 were less likely to only have one CD4 cell count; c) the proportion of patients with AIDS included in the integrated dataset was low. Even after 1995, individuals who had died were less likely to be included in the integrated dataset than those who had not and particularly individuals who died shortly after diagnosis. Their short survival was likely to have reduced their likelihood of being reported to SOPHID. An AIDS diagnosis was associated with non-inclusion before 1995 but not afterwards, probably due to the high risk of death due to AIDS before 1995, and the relatively good survival afterwards with effective treatment.

The overall effect of differential inclusion in the integrated dataset on subsequent analyses would likely be the significant over-estimation of positive outcomes and under-estimation of negative outcomes. This is primarily because individuals who had died were less likely to be included than others, underestimating late diagnosis and consequent mortality (many of those who died shortly after diagnosis without a CD4 cell count were probably severely immunocompromised), underestimating the incidence of AIDS and death, and overestimating rates of CD4 increases after initiation of ART. The differential inclusion of population groups probably had a similar effect on overall outcomes because the adjusted odds ratios were marked, indicating that certain groups were substantially under-represented. As the under-represented groups were those also likely to have poorer outcomes, due to either socioeconomic status (e.g. non-white ethnicities and heterosexuals) or biological mechanisms (e.g. older aged individuals subject to faster progression of infection).

Greater representation in the integrated dataset may be due to a combination of longer follow-up and less emigration, and therefore increasing information and opportunity to link records. Coding errors of less common surnames may also be a contributing factor among individuals of non-white ethnicities. Bias towards these groups and towards London may also be due to higher completeness and consistency of reporting from large urban clinics with electronic databases and data managers and the earlier, and therefore more complete, collection of local patient ID numbers from London. The greater record linkage of records since 2000 was probably due to the greater completion of local patient ID numbers on reports of new HIV diagnoses by clinicians.

The formation of the integrated dataset also resulted in the exclusion of records for the majority of patients without ethnicity or risk group reported. These individuals were likely to have poorer outcomes causing subsequent population summary outcomes to be over/under-estimated but relatively little in magnitude due to the relatively small numbers in these groups. The additional exclusion of the few remaining individuals with no reported risk group or ethnicity disproportionately excluded deaths but this created probably only a marginal additional bias. Similarly, there was probably little effect of excluding the small number of children infected vertically from subsequent analyses.

There was also evidence of inclusion bias arising from record linkage due to the delay in measuring CD4 cell counts. Individuals diagnosed between 2005 and 2007 were less likely to be included in the integrated dataset than individuals diagnosed between 2000 and 2004. This is likely to be largely due to individuals diagnosed in 2007 who had their first CD4 cell count in 2008 or later.

The inclusion bias introduced due to formation of an integrated dataset is likely to be a particular issue for analyses based on linked electronic patient records collected through complementary surveillance systems but may similarly affect cohort studies based on the integration of data from various sources[249;260-262]. This could be addressed analytically by sampling or weighting to adjust for greater representativeness. For example, inverse probability weighting is a relatively simple method to proportionally account for missing observations according to the number of records missing from each category when the population from which the sample comes from, and its breakdown by category,

are known. HIV surveillance data may not capture the whole HIV-diagnosed population, but are very comprehensive and nationally representative, and therefore provide a good opportunity to identify, quantify and account for bias. However, complex bootstrapping techniques are required subsequent to the inverse probability weighting to calculate confidence intervals. Furthermore, inverse probability weighting assumes that individuals are missing at random and therefore that individuals included in the analysis have the same risks/outcomes as those not included with the same characteristics. A neater solution for prospective data would be to minimise the source of this bias by integration to create a single HIV surveillance system or extensive follow-up to ensure near-complete record linkage. It is important to note that results of subsequent analyses may be easier to interpret in light of the systematic inclusion bias described above (where set algorithms determine the record linkage) than if the records had been linked manually and subjectively.

5.5.2 Potential bias due to data being right-censored

These results showed that potential bias due to data being censored had to be considered in subsequent analyses. There was evidence of overrepresentation of individuals with CD4 cell counts within 91 days of HIV diagnosis, which was even more marked with a cut-off of 182 days. This was partly due to the formation of the integrated dataset, requiring record linkage to CD4 Surveillance, which resulted in all individuals included who were diagnosed in 2007 having CD4 cell counts by the end of 2007.

AIDS diagnoses and deaths were more likely to be preceded by a CD4 cell count within 365 days if the last CD4 cell count before the event occurred in 2005 or 2006. Because data were censored at the end of 2007, the proportion of individuals with CD4 counts within the previous 365 days of AIDS or death was probably overestimated for these years. Furthermore, all individuals who had CD4 counts in 2007 and who had died or been diagnosed with AIDS were excluded from analyses of time from previous CD4 cell count because these events cannot have occurred more than 365 days after the CD4 cell count. This would be expected to bias the analysis of the incidence of AIDS and death (chapter 7)[247]. This was also evident in the high rate of permanent LTFU in 2005 and 2006, which was likely to be an artefact of incomplete follow-up: almost 5,000 individuals had their last CD4 cell count in 2005 or 2006 and were not known to have died within 365 days but individuals with subsequent CD4 cell counts reported after the end of 2007 would have been misclassified.

5.5.3 Potential bias due to early mortality (left-censoring)

Deaths shortly after HIV diagnosis accounted for a substantial proportion of all deaths among those who were not included in the integrated dataset, particularly among those diagnosed since 1995. Additionally, individuals who died shortly after HIV diagnosis were less likely to be included in the integrated dataset than individuals who died more than a month after HIV diagnosis. This was probably due to a combination of lack of opportunity to measure CD4 cell counts before death and under-reporting of inpatient care to SOPHID. This measure of bias is very difficult to ascertain in cohort studies and very difficult to

correct for because these individuals are not included in the dataset[247]. Surveillance data provides the advantage of quantifying the inclusion bias and demonstrates the importance of maximising the completeness of mortality data from national death statistics[257;263].

The low inclusion of records for individuals who had died was likely to result in an under-estimation of the incidence of death, particularly because individuals who died shortly after HIV diagnosis would contribute little follow-up time. However, this was less likely to affect the pattern of incidence between groups.

5.5.4 Potential bias due to differential loss to follow-up

5.5.4.1 Permanent LTFU (emigration, under-reported deaths or inability to link records)

Younger individuals were more likely to be permanently LTFU than older individuals and also more likely to have only one CD4 cell count. This was also the case for Black Africans, in particular, but also other non-white individuals compared to white individuals. This may reflect temporary migration to the UK of foreign students, business people and unsuccessful asylum seekers[264;265]. Young people are also likely to be more mobile, and more likely to marry, resulting in higher apparent LTFU due to the reporting of different patient identifiers. Permanent LTFU may also reflect avoidance of healthcare systems by migrants (including unsuccessful asylum seekers and undocumented migrants) who fear being forcibly relocated, detained or deported[265-269]. Individuals choosing not to remain where they are forcibly relocated and subsequent avoidance of healthcare systems may partly explain the higher

permanent LTFU among individuals from outside London. Youth[256;270-278] and race[265;272;277;279-281] have been widely found to be associated with LTFU with the latter a likely proxy for foreigner/migrant status[273;275;276] or eligibility for care[274].

Rates of permanent LTFU were lower for MSM than for heterosexuals and particularly high for IDU after controlling for ethnicity and similar differences were seen for the odds of having only one CD4 cell count. This may also reflect the higher emigration of individuals who were not born in the UK, who may be more prevalent among heterosexuals and IDU than among MSM. Many HIV-infected IDU in the UK were born and infected in southern Europe, many HIV-infected white heterosexuals were born and infected in Africa whereas many HIV-infected white MSM were born and infected in the UK[265;282]. However, there may also be different emigration rates for individuals who were born in the UK because MSM may be more likely to stay in the UK due to criminalisation of homosexuality in some other countries[283]. Additionally, the complex lifestyles and marginalisation of IDU, due to the illegality of drug use, are also likely to account for a substantial degree of permanent LTFU[284]. MSM may be more likely to be reported with consistent identifiers as they may be less mobile, due to the attraction of living in areas with a high density of other gay men and the concentrated location of these areas, and less likely to change names due to marriage than other individuals[285]. High rates of LTFU among IDU have been well documented[270;273;276;281;286]. However, other studies did not find higher rates among heterosexuals than MSM, which may be due to inclusion and adjustment for migrant status in those studies[273;276].

Rates of permanent LTFU were greater for individuals with low CD4 cell counts and particularly substantial for those with CD4 cell counts less than 100 cells/mm^3. This may have been due to a combination of under-reported deaths and the repatriation of people born outside the UK who didn't expect to live very long. The under-reporting of deaths was also indicated by the relatively high rate of permanent LTFU in 1995/6 at a time when the number of deaths among HIV-infected individuals in E,W&NI was at a maximum. In contrast, there were lower odds of permanent LTFU after an AIDS diagnosis, which may indicate that people who had been ill were less likely to emigrate and that there was more complete reporting of deaths after an AIDS diagnosis than other deaths (reflecting the under-reporting of non-AIDS deaths). Similar results were found in other studies of permanent LTFU, which noted the limitations of under-reported deaths and the potential for the transfer of care to other clinics within or outside the country[250;256;265;270;273;276]. However, the median CD4 cell counts prior to permanent LTFU were greater than those prior to temporary LTFU, indicating that it was unlikely that permanent LTFU was mostly due to death.

These findings may indicate bias in the overall estimates of short-term mortality, incidence rates of AIDS and death, and rates of change of CD4 cell counts after starting ART due to the under-representation of events. If individuals with greater permanent LTFU were also more likely to die, develop AIDS or have faster CD4 cell count responses to ART, then the overall results of subsequent analyses would be under-estimated. Additionally, differential LTFU between groups may bias their comparative event rates in subsequent analyses particularly due to fewer events rather than reduced follow-up time.

The overall incidence rate of permanent LTFU in the integrated dataset was 5.6 (10,365/185,768) per 100 PYFU and was 4.1 (4,508/108,832) per 100 PYFU between 1997 and 2004. This was very similar to the incidence of 3.7 per 100 PYFU found in a multi-centre study of combined clinical cohort data of 12,304 patients from mostly European countries and 4.3 per 100 PYFU found in a multi-centre study of combined clinical cohort data of 1,756 patients from a region in France[270;276]. These rates were similar, which suggested that the limitations of record linkage between national HIV surveillance data were offset by the benefits of national coverage to account for transfers between clinics. However, the bias introduced by each of these reasons for LTFU may be different and there is some evidence that individuals who transfer their care may be seeking expert care and therefore over-represented in centres of excellence that are often used for cohort studies[287]. Rates were lower (3 per 100 PYFU) in an American cohort study but this was an interval cohort study of enrolled patients and therefore these patients were more likely to be attendees most engaged with their clinical care[249;288;289]. Rates were substantially higher (17.2 per 100 PYFU) in less developed countries despite the similar determining factors[275;290]. Many studies could not provide rates for comparison because rates were not measured. Furthermore, factors associated with LTFU were often not determined because numbers were relatively small.

It is important to note that rates of LTFU from clinical cohorts may be much higher than in the integrated dataset due to undocumented transfers to care elsewhere. In a collaborative study with a large HIV clinic in south-east London, we showed that almost two-fifths of patients were permanently LTFU from the

clinic but that almost half of these subsequently received care at another clinic in the UK (through record linkage to the national SOPHID dataset)[265].

*5.5.4.2 **Temporary LTFU** (non-attendance, under-reporting or inability to link records)*

Rates of temporary LTFU decreased with age, which may reflect knowledge of the lower risk of progression among young people, as shown by other studies [275;279;280;291]. Temporary LTFU was higher for MSM than for recipients of blood/blood products, but lower than for other groups, and particularly high for IDU. Similar factors were found to be associated with temporary LTFU in the literature[279;281;292]. Recipients of blood/blood products were likely to be closely associated with healthcare as many were haemophiliacs. MSM were likely to have high awareness of the need for regular monitoring of their infection whereas heterosexual women diagnosed antenatally were likely to have frequent check-ups during pregnancy and be associated with healthcare through care of their children[293]. IDU were least likely to prioritise their health and have complex lifestyles resulting in irregular attendance with services[284]. Rates of temporary LTFU were high for black individuals and were particularly high for black Caribbeans compared to white individuals. This was probably due to complex housing, employment, immigration and financial needs, associated with deprivation, which often have to be prioritised above long-term healthcare needs[267;294-297]. Stigma and discrimination of HIV infection are also substantial among black communities, which can discourage disclosure and complicate access to healthcare[294;295;297]. These factors all form personal/cultural,

structural, and financial barriers to care[298]. Higher rates of LTFU among minority ethnic groups have been reported in the literature[279-281;286;291;298].

Rates of temporary LTFU were high at both low and high CD4 cell counts. Temporary LTFU at high CD4 counts has been reported previously and may be due to the lower risk and less need for regular monitoring[275]. Other studies that did not find an association between CD4 cell counts and temporary LTFU only had the power to compare high and low CD4 counts[291;298]. CD4 cell counts were monitored more frequently after low CD4 counts and there were lower rates of temporary LTFU after an AIDS diagnosis indicating the close monitoring of infections at times of high risk of mortality, consistent with the literature[279;291]. The high rates of temporary LTFU at low CD4 cell counts do not seem to fit with these results unless these patients were admitted to hospital for inpatient care and the CD4 counts during this time were reported with different patient identifiers and therefore not linked to previous records.

Temporary LTFU was likely to bias overall results from subsequent analyses towards over-estimation in contrast to permanent LTFU. This was due to the under-estimation of follow-up time but had minimal impact on the number of events. This was because deaths could not have occurred during the period of temporary LTFU and AIDS diagnoses were unlikely to be missed, unless due to non-reporting, because individuals with AIDS were likely to present for healthcare and be cared for by an HIV specialist.

Comparisons of the overall incidence of temporary LTFU were difficult to make with studies reported in the literature. Many studies that considered missed appointments or retention in studies compared proportions and did not formally calculate PYFU or incidence rates. A study of a prospective cohort, which divided the number interviewed in the last 12 months by the number enrolled, found retention rates of 79 to 95% in different sites[280]. However, as other studies[279;281], this did not separate temporary and permanent LTFU. A study that interviewed women about their health service use found that 92% had visited an outpatient or emergency department in the last six months[286].

5.5.4.3 Summary of LTFU analyses

Formation of the integrated dataset from national surveillance data and national mortality statistics was believed to have minimised the under-reporting of deaths, particularly those that occur at home or in a hospice, which is a significant factor for clinic cohorts. However, relatively high rates of permanent LTFU after low CD4 cell counts, among IDUs and in 1995/1996 suggest that not all HIV surveillance records were linked to their appropriate death reports. National follow-up through linkage of records from almost all immunology/haematology laboratories minimised the over-estimation of LTFU due to transfer between clinics within E,W&NI. However, the inherent disadvantage of analysis of LTFU from the integrated dataset was the potential for over-estimating rates due to the inability to link patient records over time. Yet, there was no plausible reason for changed patient identifiers to be associated with CD4 cell counts or a previous AIDS diagnosis, or so closely aligned with results from other studies with regard age, risk and ethnicity.

Therefore, the differential LTFU found in the integrated dataset, except for changes over time, was likely to reflect true LTFU and needed to be considered as a source of potential bias in subsequent analyses.

5.5.5 Potential bias due to the differential inclusion of events: differential event rate in Poisson analysis

5.5.5.1 Time from HIV diagnosis to first CD4 cell count

The proportion of individuals with a CD4 cell count within 91 days of their HIV diagnosis increased significantly over time suggesting that the CD4 Surveillance data became more complete, record linkage improved, or that there was a true reduction in the time between diagnosis and measurement of CD4 cell counts (the particularly high proportion with CD4 cell counts within 91 days of diagnosis in 2007 was probably an artefact created by censoring the data at the end of 2007 (Section 5.4.4.1)). A lower proportion of younger people had a CD4 count within 91 days of HIV diagnosis than older people, which may reflect knowledge of the greater risk of progression for older people or greater concerns about health among older people[124;299]. MSM were less likely to have CD4 counts within 91 days of HIV diagnosis compared to women diagnosed antenally but more likely than other heterosexuals and IDU. IDU were least likely to have timely CD4 counts reflecting their complex lifestyles resulting in irregular attendance with services[284]. Interestingly, heterosexual women diagnosed antenally were less likely than MSM to have CD4 cell counts within 30 days of diagnosis suggesting that the greater return for follow-up care within 91 days may be partly determined by the antenatal care schedule. The earlier CD4 counts for MSM than for other heterosexuals may be due to greater knowledge

of HIV among MSM and less difficulty in accepting a diagnosis. White individuals were more likely to have a CD4 count within 91 days of HIV diagnosis compared to black African and black Caribbean individuals, which may also relate to awareness of, and stigma associated with, HIV infection as well as lower prioritisation of health needs due to other economic concerns[267;294-297]. AIDS at the time of HIV diagnosis was associated with more immediate measurement of CD4 cell counts, reflecting the urgent need for information to help manage the infection due to the high risk of death[124;299].

Many of these results were consistent with those of other studies. Previous publications reported delayed initiation of care associated with calendar time[300-302], age[300;303;304], IDU risk[167;301;305;306], non-white ethnicity[169;304;305;307], and a lack of AIDS symptoms or high CD4 cell count[308;309]. The difference between MSM and heterosexuals was not found in the literature but this may be due to lack of power. However, studies did find that individuals who were not aware of their HIV risk before testing or who had no previous HIV test experienced delayed initiation of care[167;306]. Additionally, qualitative studies found that women's psychological responses to learning of their diagnosis were most strongly associated with delayed care seeking, while for MSM, this reflected a strong reluctance to begin ART[310;311]. Interestingly, women diagnosed antenatally were reported to have long delays before entry into care in two studies in contrast to results from the integrated dataset, which may reflect different pathways into, and eligibility for, care between the Americas and the UK[312;313]. One London study found similar proportions of black African and white individuals received HIV care within one month of HIV diagnosis but did not control for other

factors[171]. In addition to the factors investigated from the integrated dataset, socioeconomic factors, such as unemployment[306] and lack of health insurance[169;302;309], and healthcare factors, such as lack of personal post-test discussion or counselling[167;170;306], anonymous and first-time testing[314] and diagnosis at an unfamiliar testing site[169;170;305] were associated with delayed access to care.

Overall comparisons with time to first CD4 cell count or initiation of HIV care were hampered by the variety of definitions used, the study characteristics and changes over calendar time. Delayed entry to HIV care was reported for 17-58% of individuals between 1993 and 2003 and largely from the United States, where pathways into, and eligibility for, care may differ significantly from the UK[167;169;302;303;305;306;313;314]. However, these studies suggest that the evidence of delays between HIV diagnosis and first CD4 cell count demonstrated from analysis of the integrated dataset were true delays of entry into HIV care rather than artefacts of linked HIV surveillance data.

Analysis of time from HIV diagnosis to first CD4 cell count quantified the extent to which the subsequent analysis of late diagnosis was biased by exclusion of records that did not have a CD4 count within a defined time. Over one in seven individuals diagnosed since 1995 did not have a CD4 count within the following 91 days, and this would have resulted in inclusion bias if the CD4 counts at the time of HIV diagnosis were different for these individuals than those included in the analysis. Extending the cut-off time from HIV diagnosis would increase the proportion with CD4 counts included in the analysis but CD4 counts further from

the date of HIV diagnosis would be less representative of the true CD4 cell count at the time of HIV diagnosis and potentially subject to treatment or other effects. The overall estimate of late diagnosis would be affected by the differential inclusion of groups of individuals due to the different proportions with CD4 cell counts within the cut-off period from HIV diagnosis. For example, individuals with AIDS at HIV diagnosis would be over-represented and if their CD4 counts at the time of HIV diagnosis were lower than for other individuals, the overall estimate of late diagnosis would be over-estimated. Additionally, differential inclusion could result in biased comparisons between groups if those included were not representative of all individuals in the group.

5.5.5.2 Time from CD4 cell counts to AIDS and death

The proportion of individuals who had their last CD4 cell count measurement within 365 days of death was higher for those with CD4 counts less than 50 cells/mm^3 and for those with a previous AIDS diagnosis. This was likely to reflect the increased risk of death at these low CD4 counts and after AIDS[124;299] but also may reflect the increased frequency of CD4 count monitoring at low CD4 cell counts and after AIDS (Section 5.4.3.2). A similar effect was observed for CD4 cell counts within 365 days of AIDS for those with CD4 counts less than 150 cells/mm^3 for similar reasons. There was some evidence of a lower proportion with CD4 count within 365 days of death among younger people and strong evidence with time to first AIDS diagnosis, which also was likely to be due to slower progression of infection among younger individuals[124;299]. Only IDU were less likely than MSM to have a CD4 cell count within 365 days of AIDS or death, which probably reflects their less frequent attendance for health

services because IDU have higher rates of AIDS and death[124;299]. Individuals last CD4 cell counts were less likely to be within 365 days of death if they were measured in 1995/6. If this was due to no CD4 count being measured during that time (rather than unreported or unlinked data), this may reflect less benefit derived from monitoring levels of immunosuppression in the pre-HAART era when fewer treatment options were available to patients with low CD4 counts. There was no significant effect of ethnicity on the odds of having a CD4 count within 365 days of AIDS or death despite some ethnic groups having CD4 counts monitored less frequently and higher rates of temporary and permanent LTFU than others.

It is important to note that less than a third of individuals with AIDS diagnoses within three months of HIV diagnosis had a previous CD4 cell count compared to nine-tenths where AIDS was diagnosed more than three months after HIV diagnosis. Exclusion of follow-up time before the first CD4 cell count in analysis of incidence estimates was therefore a likely source of bias.

Analysis of the time between CD4 cell counts and AIDS and death, together with the analyses of temporary and permanent LTFU, provided some measure of the extent to which the analysis of incidence rates of AIDS and death were biased. One in ten deaths and almost one in six first AIDS diagnoses occurred more than 365 days after the last CD4 count and were therefore considered to have occurred during a period of LTFU. These events and time could be included in the analysis by extending the cut-off but the CD4 cell count, and other time-updated variables, would be increasingly less representative of the

true level of immunosuppression at the time of death. Therefore right-censored follow-up time was necessary to limit misclassification bias[247]. However, excluding these periods of LTFU was likely to bias the incidence of events towards under-estimation because LTFU was probably due to death in some cases and was higher at low CD4 cell counts and after an AIDS diagnosis.

Further work could provide greater insight to the relationships between CD4 cell count monitoring and deaths by investigating causes of death in the HARS data but would have to consider the particular biases in death reports (Section 2.2.8).

5.5.6 Potential bias due to the differential monitoring CD4 cell counts: differential contribution to analysis of rates of change of CD4 cell counts after starting ART

CD4 cell counts were monitored more frequently at lower CD4 cell counts and in individuals with a previous AIDS diagnosis, particularly at the time of the AIDS diagnosis. Individuals with a low first CD4 cell count were also more likely to have more than one measurement. These results reflect the need for information about the level of immunosuppression, which is required to manage HIV infection at times when there is a high risk of death[124;299;315]. For individuals with a reported date of starting ART, there was also an increased frequency of CD4 cell count measurements at the time of starting ART as recommended by guidelines[123]. More intensive monitoring is recommended to confirm the need to start ART and to determine the immunological response[233]. In particular, national guidelines recommend CD4 measurement by GUM/HIV specialists,

and not necessarily at the time of HIV diagnosis, which may lead to a variety of practice and delays for some individuals who need prompt ART[105;316].

There were differences in the frequency, number and duration of monitoring of CD4 counts between groups and over time but it is likely that some of these effects were confounded and would not demonstrate statistical significance in multivariable analysis. A key potential confounder of these effects is calendar time, because CD4 counts were measured less frequently since the mid-1990s. This may reflect real trends in monitoring due to the reduced risk of progression with effective ART but would affect the risk groups and ethnic groups differently. Differences in the frequency and number of CD4 counts may introduce bias to analyses of rates of change of CD4 counts after starting ART[248].

There were few patients with only a few CD4 cell counts over a long period of time and few with many measurements over a short period. The former may be due to inability to link all CD4 counts for an individual (Chapters 3 and 4), under-reporting of CD4 counts (Section 2.3.2), right censoring (Section 5.5.2), or long-term LTFU (Section 5.5.4). Right-censoring is particularly likely to explain the high proportion of patients with only one CD4 count that were diagnosed between 2005 and 2007 because an increasing proportion of these individuals had limited time to return for subsequent CD4 measurements. Death in the same or subsequent year as the CD4 count was taken into account in the analysis but other deaths and unreported deaths were not. Long-term LTFU (either a loss to clinical follow-up or a limited ability to link records over time) was likely to account for the majority of the remaining patients with only one

CD4 count because this was associated with younger age, residence in London, heterosexual risk, non-white ethnicity and CD4 counts greater than 350 cells per mm^3. These factors in turn, are often associated with mobility, particularly international migrancy and dissociation from healthcare services due to uncertain legal migration status. These characteristics are important to identify because subsequent analyses may be biased by their under-representation. For example, a CD4 count close to diagnosis is required for analysis of late diagnosis, CD4 counts are used as the basis for analysis time to determine the incidence of AIDS and death, and individuals with single CD4 counts will contribute little to the analysis of CD4 count changes after initiation of ART.

Patients with many CD4 measurements over a short period may be due to inappropriate deduplication of CD4 Surveillance records or particular needs/demands for frequent CD4 count monitoring (such as virological failure or adherence monitoring). Although these reasons could not be differentiated in these analyses, the numbers were small enough to indicate that these records were unlikely to substantially bias the results of subsequent analyses.

Measurement of CD4 cell counts was more frequent in the period shortly after HIV diagnosis, irrespective of the CD4 cell count category, which probably reflects uncertainty in a single measurement and the clinical need for confirmation with a second sample[243;317]. This effect was likely have an impact on determination of baseline CD4 cell counts at the time of HIV diagnosis and was therefore taken into account in their definition (Section 6.3.1).

Guidelines before 2008 did not state how often CD4 cell counts should be monitored or how often patients should attend as there was little evidence for the historical practice of three-monthly follow-up[258]. Modelling suggests that monitoring should vary according to the status of infection, which may reflect actual practice although local policies vary[318]. These results could be used to inform guidelines for best practice after validation for cost-effectiveness.

Association between groups of individuals or time-updated characteristics with an increased frequency or number of CD4 cell counts could result in over-representation in analyses of rates of change of CD4 counts after starting ART. If these groups or characteristics were also associated with rates of change of CD4 counts after starting ART then this would bias those results. These associations were therefore considered in more detail in chapter 9 once dates of starting treatment had been considered.

5.6 Conclusion

Surveillance data for public health action may not be as accurate as clinical data or data collected and validated for research studies. The greater number of individuals included limits the manual consistency checking of integrated data at the individual level that is possible for smaller datasets, and therefore increases the risk of errors. These errors potentially cause outliers in analyses such as very short or long times between CD4 cell counts and large fluctuations between CD4 cell counts. The collection of data via multiple sources and the need for subsequent centralised record linkage, require centralised resources and limit the completeness of integrated datasets. However, this approach

minimises the resources required at local level and some errors can be removed with little concern for reducing the high power of statistical analyses due to the large numbers involved. Furthermore, triangulation allows direct quantification of inclusion bias as the disaggregate characteristics of the denominator are known. Surveillance that collates data already collected at a local level is therefore cost-effective[319], has full national coverage and therefore large numbers (allowing monitoring of rare events and minority groups), and is not time-limited (therefore timely, continually updated, and relevant for public health action). Bias is minimised because UK HIV surveillance data are collected prospectively and therefore cannot be biased by knowledge of outcomes. However, it is important to note that many statistical tests are performed on these data, both within this thesis and within wider surveillance work, and therefore, some significant associations that are found could be due to chance.

Drawbacks of the integrated dataset were that the limited dataset may not have included other confounding factors and that changes in risk factors over time were not recorded. However, risk factors analysed for outcomes such as late diagnosis, the incidence of AIDS and death and rates of change of CD4 cell counts after starting ART could also be considered as proxy measures for other factors, such as socioeconomic status, HIV awareness, and adherence to therapy. However, the risk factors of risk group and ethnicity are of practical use for targeting more qualitative research about underlying reasons and for targeting prevention efforts. Misclassification bias due to errors in the measurement of risk factors, and particularly CD4 cell count measurements,

would be expected to reduce differences in the results and may have affected these analyses of surveillance data to a greater extent than clinical cohorts because of unspecific record linkage.

The analyses of temporary and permanent LTFU show that surveillance data can be used to create a pseudo-cohort with very large numbers of patients (almost 46,000) and PYFU (over 220,000). The use of CD4 cell counts to define person-years ensures that patients were under follow-up, which may otherwise be a limitation of using surveillance data. The CD4 Surveillance procedures require annual attempts to link together all CD4 counts for the same patient and account for transfers between clinics even if the CD4 cell counts were reported from different laboratories. However, imperfect record linkage within and between surveillance systems due to changed patient identifiers can result in gaps in follow-up. While considerable effort is made to minimise the linkage of records from different people and the inaccurate recording of information, surveillance data cannot be quality assessed to the same extent as data from local cohort studies. However, local cohort studies need to censor follow-up when transfers are made to other clinics or estimates of LTFU may substantially reflect transfer rates. This would be difficult in the UK due to the open access policy of HIV services. Therefore estimates of LTFU from national surveillance may be less biased by transfers than local cohort studies but more subject to bias from incomplete record linkage. Permanent LTFU from the entire UK cohort could only be explained by emigration, unreported death or death reported with different identifiers, continued care in the UK that was not reported or reported with different identifiers or complete dissociation from HIV care.

Unreported deaths were believed to be minimal due to record linkage to death notifications and under-reported continued care in the UK was also believed to be minimal due to the near complete coverage of surveillance systems. Substantial effort was made to maximise reliable record linkage and permanent dissociation from HIV care in the UK is unlikely for long periods of time because of the free availability of treatment, the open access to anonymous and confidential care, and the progressive nature of HIV infection. In contrast, the large proportion of individuals diagnosed with HIV infection in the UK who were born abroad supports the conclusion that the majority of permanent LTFU was due to emigration.

These analyses have shown that subsequent analyses must consider left and right censoring. Follow-up should start from the time of HIV diagnosis rather than at the time of the first CD4 cell count to maximise the inclusion of the numerous events that occur soon after diagnosis. However, analyses of the integrated dataset could not include individuals, follow-up time or events where LTFU occurred before a CD4 cell count was measured or where no CD4 cell counts for an individual were reported. Consideration of right censoring at the end of the study period and during periods of LTFU was important because LTFU did not occur randomly. LTFU was associated with factors that were independently associated with progression and therefore censoring was informative. Methodological strategies could therefore bias follow-up time and event rates. In particular, individuals diagnosed in 2007 should be excluded from subsequent cohort analyses due to insufficient follow-up time before censoring at the end of December 2007.

The results of these analyses indicate that 91 days between HIV diagnosis and first CD4 cell count could be an appropriate cut-off for the definition of CD4 cell counts at the time of HIV diagnosis. Additionally, a cut-off of 365 days could be appropriate to capture most variation in follow-up and therefore to minimise bias due to excluding AIDS or death events more than 365 days after a CD4 cell count measurement. These values were therefore used in subsequent analyses with sensitivity analyses to determine the influence of varying the methodology. The understanding of the data gained through these exploratory investigations would be invaluable for the interpretation of subsequent analyses.

Chapter 6. Late HIV diagnosis and consequent short-term mortality

6.1 Introduction

The estimated prevalence of HIV infection and numbers of new HIV diagnoses in the UK increased between 1998 and 2007 but a substantial number and proportion of HIV-infected individuals remained undiagnosed[223;282] and many individuals were diagnosed at a late stage of infection[80;300;320-326]. Prior to diagnosis, individuals cannot receive appropriate care, inform their sexual partners or be guided in safer sexual behaviour in knowledge of their status[327-329]. Individuals who remain unaware of their HIV status for a long time and are diagnosed late lose the opportunity to start ART appropriately. Late initiation of ART has been shown to be associated with increased morbidity and mortality[124;125;330]. Delays in diagnosis also reflect lost opportunities to reduce: i) HIV transmission, ii) the risk of death between diagnosis and initiation of HAART, and iii) costs of care[331;332]. This represented a continuing challenge for HIV testing promotion and policy-making as highlighted in the 2004 annual report of the Chief Medical Officer for England in the section "No time to wait: The importance of early diagnosis of HIV" [333]. The primary aim of this chapter was to use the integrated dataset to investigate late diagnosis and consequent mortality. Largely due to the work contributing to this thesis, late diagnosis has since been used as a measure of the effect of the Department of Health's National Strategy for Sexual Health and HIV for England, which aimed to reduce the number of undiagnosed HIV infections in the population and to reduce HIV transmission by prioritising HIV testing[111;241].

CD4 cell counts at the time of HIV diagnosis provide an indication of time from infection and of the risk of progression unaffected by ART[32;124;299;334]. The time taken from infection to diagnosis has implications for HIV prevention efforts and HIV testing strategies, and the individual and population risks of AIDS and death are crucial for clinical decision-making.

Numerous definitions have been used to define late HIV diagnosis in the literature and have generally been based either on CD4 cell counts or a clinical diagnosis of AIDS at the time of HIV diagnosis[335-342]. Low CD4 counts or AIDS generally reflect a long time from infection due to the slow rate of progression of HIV-related immunosuppression (Section 1.3.1). An uncommon exception is that individuals can experience substantial immunosuppression and also AIDS-defining illnesses shortly after infection[340;343].

Studies that have considered first AIDS diagnoses at the time of HIV diagnosis, as a proportion of all first AIDS diagnoses, do provide an indication of late HIV diagnosis because this proportion would be low if all infections were diagnosed promptly, when there is a low risk of AIDS[336;344-349]. These studies have been facilitated by the availability of surveillance data on AIDS diagnoses[185;347-349]. However, the denominator is also dependent on the availability of effective ART and prophylaxis for opportunistic infections. Effective prevention of AIDS among previously diagnosed individuals will decrease the denominator and increase the proportion of AIDS diagnosed at the time of HIV diagnosis irrespective of the time between infection and HIV diagnosis. This limits the implications that can be drawn from comparisons between countries and over time.

Additionally, studies that have considered AIDS diagnoses that occur around the time of an HIV diagnosis, as a proportion of all HIV diagnoses, also provide an indication of late HIV diagnosis[168;350-354]. Results from these studies are not affected by treatment or prophylaxis. However, the numerator is unlikely to be comparable between countries and over time because the background rates of AIDS-defining illnesses can vary between populations and over time and can occur at different levels of immunosuppression[355-359].

In contrast, low CD4 cell counts at the time of HIV diagnosis are not affected by treatment or prophylaxis and are not affected by varying AIDS incidence. The risk of AIDS or death at various CD4 counts is of high importance to the patient and physician and largely determines the clinical management of infection. Since the widespread availability of combination ART in 1996, CD4 cell counts have been used as the most reliable predictor of the short-term risk of AIDS or death to determine the need for ART (based on results from cohort studies) [124;125;299;360]. Studies have shown that there is a low short-term risk of progression to AIDS for untreated individuals with CD4 counts above 200 cells/mm^3 although viral load and age independently increase that risk[299]. Patients with CD4 counts below 200 cells/mm^3 when starting HAART have a substantially higher risk of progression to AIDS and death[124;361]. Therefore, British treatment guidelines between 1997 and 2007 have recommended that all individuals initiate HAART before CD4 counts fall below 200 cells/mm^3 but that treatment is deferred for most until counts are below 350 cells/mm^3 [129-135;362].

The ART Cohort Collaboration, which includes 13 cohort studies from Europe and North America, showed that after starting HAART the rate of progression to AIDS or death fell markedly, with most of the reduction in the first six months of therapy[124]. It is important, therefore, that individuals have the optimal chance of surviving the first six months of HAART.

Short-term mortality quantifies the implications of late diagnosis by considering it as a risk factor for death within a year of diagnosis. It is possible that late diagnosis has a long-lasting effect on survival but short-term survival requires shorter follow-up and is therefore timelier and less biased by loss to follow-up.

Time of infection cannot be reliably estimated at the individual level from CD4 cell counts at the time of HIV diagnosis due to the wide variability in rates of progression but can provide an indication of the duration of infection at the population level[363].

Studies have shown that in the UK, heterosexuals have generally been diagnosed later than MSM, and that black Africans have been diagnosed later than white individuals[164;320]. However, these factors have not been considered independently. There has been no published evidence of earlier diagnosis among HIV infected individuals in the UK over time. Short-term mortality post-HIV diagnosis had not been previously reported.

6.2 Aims

a) To use the integrated dataset to explore national trends in the proportions of adults diagnosed late with HIV infection between 1995 and 2007 and factors associated with being diagnosed late using an appropriate and generalisable definition.

b) To use the integrated dataset to determine national trends in the short-term mortality (death within the year after HIV diagnosis) of adults newly diagnosed with HIV infection between 1995 and 2007, and to determine the impact of late diagnosis and other risk factors on short-term mortality.

c) To use the integrated dataset to quantify late diagnosis and consequent short-term mortality among individuals heterosexually infected with HIV and diagnosed between 2000 and 2007. To identify risk factors for late diagnosis and short-term mortality among these heterosexuals including additional HIV surveillance information, such as country of birth, which was collected and followed-up for heterosexuals diagnosed in England, Wales and Northern Ireland (Section 2.2.9) since 2000.

d) To use the integrated dataset to estimate the short-term mortality that could be prevented if all adults, except those recently arrived in the UK with low CD4 cell counts, were diagnosed promptly.

6.3 Methods

6.3.1 Data preparation for analysis of late HIV diagnosis

The main study population comprised all adults (aged 15 years or older) reported as newly diagnosed with HIV infection in England, Wales and Northern Ireland in the integrated dataset between January 1995 and December 2007. Late diagnosis was defined by CD4 cell counts and the analysis was based upon the subset of the population who had a reported CD4 cell count measurement in the 91 days following the date of HIV diagnosis, subject to sensitivity analysis. This cut-off was chosen, based on previous analyses (Sections 5.4.4.1 and 5.5.5.1), to provide the least-biased estimates of late diagnosis by maximising the number of patients included in the analysis and minimising the possibility that HAART had affected the CD4 cell counts analysed[124;364]. The CD4 cell count closest to the date of HIV diagnosis was used in the analysis. Sensitivity analysis also considered definitions using the first two CD4 cell counts to match recommendations to account for measurement variation[233]. Individuals with CD4 cell counts less than 200 cells/mm^3 within 91 days of the date of HIV diagnosis were considered late diagnosed.

Firstly, CD4 cell counts at the time of HIV diagnosis for individuals reported to have previously received care were compared to those of individuals who had no evidence of previous care at the time of HIV diagnosis. Based on these results, these records were excluded from further analysis.

Secondly, analyses were carried out to determine whether a significant proportion of individuals with CD4 cell counts less than 200 cells/mm^3 could be misclassified as diagnosed late due to seroconversion[343]. Seroconverters were considered to be individuals diagnosed with seroconversion illness or evidence of a negative HIV test within the year prior to HIV diagnosis. Seroconverters with low CD4 counts who were fast progressors could be considered to be diagnosed late in terms of their risk of mortality but not in terms of preventative mortality. Furthermore, fast progressors could not be reliably distinguished from individuals who experienced temporary immunosuppression at seroconversion. Therefore seroconverters (Section 6.4.1) were excluded from further analysis.

Risk factors considered in the analysis of late diagnosis included the calendar year of HIV diagnosis in the UK (grouped into two-year periods except for 2007, which was considered separately because these data may have been significantly affected by inclusion bias – Sections 5.4.1 and 5.5.1), age group, geographic region of HIV diagnosis, risk group and ethnic group.

Sensitivity analysis involved: a) increasing the CD4 cell count cut-off for the definition of late diagnosis to 350 cells/mm^3; b) defining late diagnosis by either AIDS within three months of an HIV diagnosis or CD4 cell counts less than 200 cells/mm^3 within 91 days; c) defining late diagnosis as either both, the mean of, or either of, the first two CD4 cell count measurements (first within 91 days of HIV diagnosis and second within 91 days of the first) being less than 200 cells/mm^3; d) using CD4 cell counts within either 30 days, 60 days or 182 days of the HIV diagnosis.

The sub-analysis of late diagnosis among HIV-infected heterosexuals analysed the first CD4 count within 91 days of HIV diagnosis between January 2000 and December 2007. The additional factor considered as a possible determinant of late diagnosis was a combined factor including more detailed information about the risk group collected through follow-up. This information was hierarchically combined as follows: probable infection from a high-risk partner (MSM, IDU or recipient of blood/blood products), probable infection abroad (by world region of infection), and, for those probably infected in the UK, probable transmission from someone infected inside or outside the UK (Appendix A.5).

6.3.2 Data preparation for analysis of short-term and preventable mortality

Short-term mortality was defined as death from any cause within a year of the date of HIV diagnosis. In this analysis, the year of HIV diagnosis was replaced by the period of HIV diagnosis, which was categorised according to the availability of ART and its effectiveness at reducing mortality [365]. The pre-HAART era was considered to be 1995 only, the introduction of protease inhibitors and availability of effective combination therapy defined the early-HAART era as 1996 and 1997, the availability of non-nucleoside reverse transcriptase inhibitors (NNRTIs) and improved use of combination therapy defined the mid-HAART era as 1998 to 2002 inclusive, whereas the introduction of other drug classes and wider choices for combination therapy defined the late-HAART era as 2003-2007. Sensitivity analyses were carried out by defining short-term mortality using different cut-off times from HIV diagnosis to death:

three months; six months and 24 months. Sensitivity analyses were also carried out using multiple categorisations of CD4 counts, the logarithm of CD4 counts, late diagnosis defined by CD4 counts less than 350 cells/mm^3, and late diagnosis defined by CD4 counts less than 200 cells/mm^3 or AIDS at diagnosis.

Preventable mortality was estimated by building simple scenario models based on proportions recently arrived in the UK, proportions diagnosed late, and the fraction of observed deaths within a year of diagnosis. Preventable deaths were assumed to be those that occurred among individuals diagnosed late who had the opportunity for earlier diagnosis as they had not recently arrived in the UK. It was assumed that individuals with a CD4 cell count measurement within 91 days of HIV diagnosis were representative of all individuals diagnosed. It was also assumed that the mortality rate could be reduced to that among individuals diagnosed promptly. The analysis was repeated assuming that heterosexuals without a reported country of birth or year of arrival were in the UK at least two years prior to HIV diagnosis.

6.3.3 Statistical analysis

The definition of a binary outcome permitted the use of multivariable logistic regression to determine adjusted odds ratios (AOR) and 95% confidence intervals (CI).

6.4 Results

6.4.1 CD4 cell counts for all adults newly diagnosed with HIV

There were 42,342 individuals newly diagnosed with HIV infection between 1995 and 2007 of whom 36,295 (85.7%) had CD4 cell counts measured within 91 days of their HIV diagnosis. The proportion of the latter diagnosed with CD4 cell counts less than 200 cells/mm^3 was 33.9% (12,291/36,295).

Just over one in nine individuals (12.0% [5,083/42,342]) had evidence of previous care of whom 22.6% (1,150) were reported to have started ART before their date of new HIV diagnosis. Most individuals (85.5% [4,344]) with evidence of previous care had CD4 cell counts within 91 days of their new HIV diagnosis date, of which 25.8% (1,120/4,344) were less than 200 cells/mm^3 compared to 35.0% (11,171/31,951) for those without evidence of previous care.

There were 2,567 individuals with evidence of recent infection (excluding those with previous care) of whom 87.9% (2,256) had CD4 counts measured within 91 days of their HIV diagnosis. Fewer (8.6% [195/2,256]) of these individuals had first CD4 cell counts less than 200 cells/mm^3 than individuals without evidence of recent infection or previous care (37.0% [10,976/29,695]) ($p<0.01$).

6.4.2 Late diagnosis of adults newly diagnosed with HIV without evidence of recent infection or previous care

There were 34,692 adults in the integrated dataset without evidence of recent infection or previous care who were diagnosed between 1995 and 2007.

Overall, 85.6% (29,695) of individuals had CD4 cell counts within 91 days of their HIV diagnosis, increasing from 61.7% (810) in 1995 to 97% (2,878) in 2007. CD4 cell counts reported to CD4 Surveillance were distributed similarly to those reported to HARS (Figure 6.1). The overall proportion diagnosed late, with CD4 counts less than 200 cells/mm^3 was 37.0% (10,976). The median CD4 count of those diagnosed late was 76 cells/mm^3 (IQR 27, 135) compared to 410 cells/mm^3 (IQR 300, 566) for those not diagnosed late. The associations between late diagnosis and date of HIV diagnosis and between late diagnosis and age at HIV diagnosis were not linear and not easily reflected by simple transformations (Figures 6.2, 6.3). Therefore date of HIV diagnosis and age were categorised for subsequent analyses.

Figure 6.1. Distribution of CD4 cell counts within 91 days of HIV diagnosis

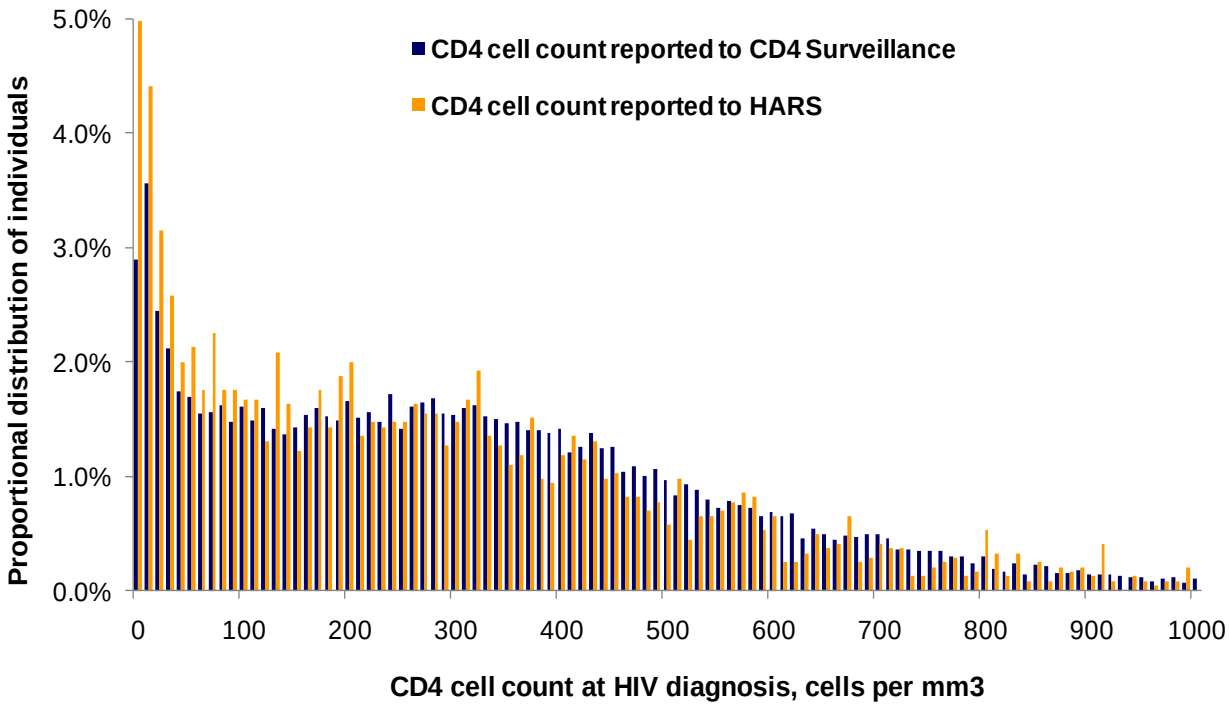


Figure 6.2. Percentage of individuals diagnosed late by year of HIV diagnosis

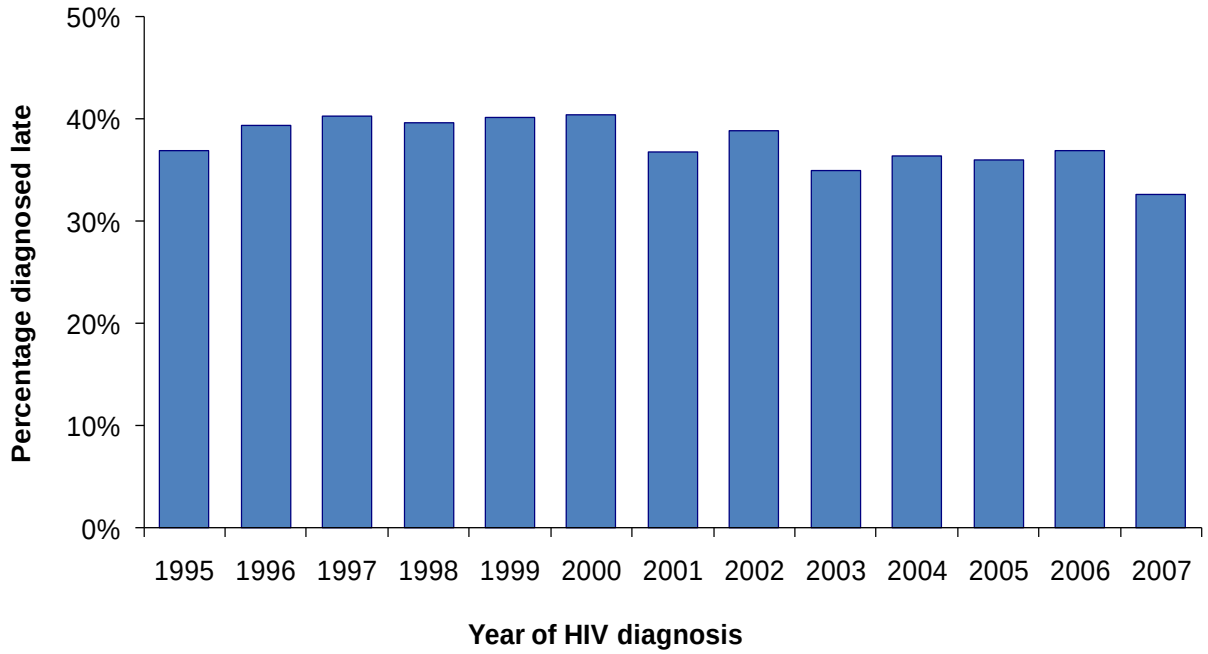
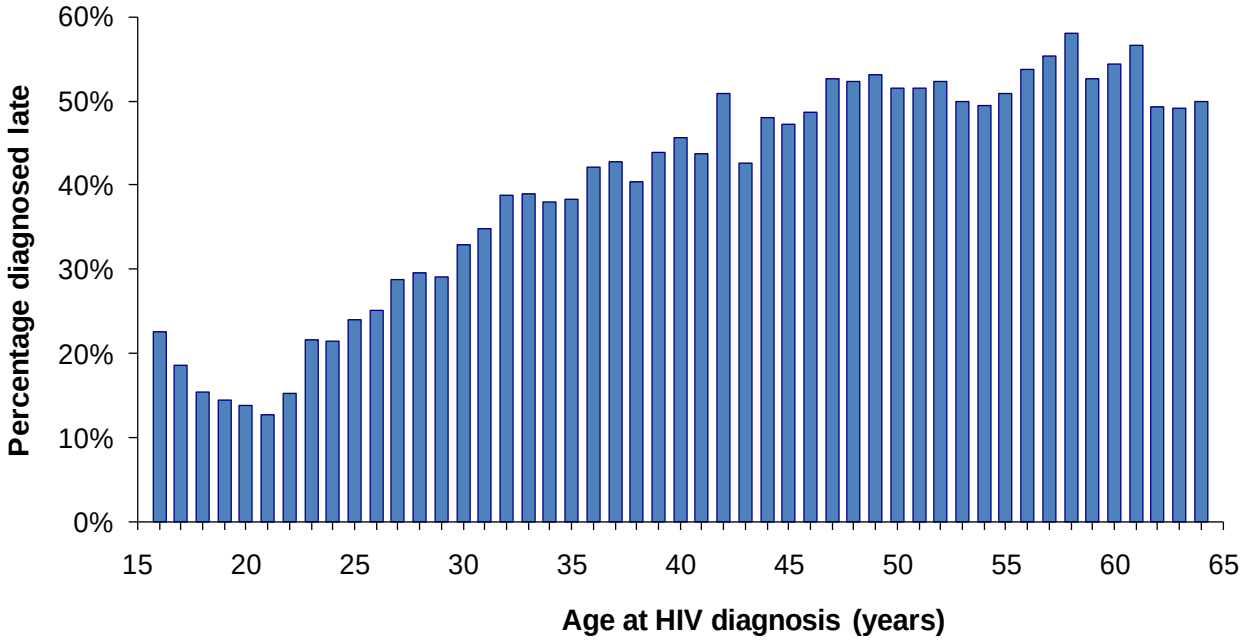


Figure 6.3. Percentage of individuals diagnosed late by age at HIV diagnosis



The overall percentage diagnosed with CD4 cell counts less than 200 cells/mm^3 (within 91 days of HIV diagnosis) declined from 40.3% (1,292/3,203) in 1999/2000 to 32.6% (938/2,878) in 2007 (Table 6.1). This decline was statistically significant and continued throughout 1995/6 to 2007. The proportion of individuals who were diagnosed late increased with increasing age, from 17.6% (593/3,362; AOR 0.32, 95% CI 0.28, 0.35) among individuals aged 15-25 years to 54.5% (675/1,239; AOR 1.87, 95% CI 1.65, 2.13) among individuals aged older than 54 years (compared to individuals aged 34-39 years). Individuals diagnosed outside London were marginally more likely to be diagnosed late (37.4% [5,177/13,827] compared to 36.5% [5,799/15,868]; AOR 1.07, 95% CI 1.02, 1.13). In comparison to MSM (28.0% [3,121/11,142]), all groups were more likely to be diagnosed late except heterosexual women diagnosed antenatally (22.0% [447/2,034]; AOR 0.80, 95% CI 0.70, 0.91). A particularly high proportion of recipients of blood/blood products were diagnosed late (59.1% [81/137]; AOR 2.54, 95% CI 1.78, 3.63 in comparison to MSM). All other ethnic groups were more likely to be diagnosed late than white individuals (although the AOR for Indian/Pakistani/Bangladeshi individuals was only marginally significant statistically).

Table 6.1. Proportion diagnosed late and odds ratios for late diagnosis – adults in the integrated dataset with CD4 counts within 91 days of their HIV diagnosis

	Number of patients with CD4 cell counts at HIV diagnosis n (% of all)	Percentage of patients diagnosed late (CD4<200 cells per mm^3) % (n)	Multivariable adjusted odds ratio for late diagnosis (95% confidence interval)
Period of HIV diagnosis			
1995-1996	1,701 (6)	38.2 (650)	**1.14 (1.01, 1.30)**
1997-1998	2,213 (7)	40.0 (886)	1.08 (0.97, 1.22)
1999-2000	3,203 (11)	40.3 (1,292)	-
2001-2002	5,466 (18)	37.9 (2,073)	**0.82 (0.75, 0.90)**
2003-2004	7,505 (25)	35.7 (2,683)	**0.72 (0.66, 0.79)**
2005-2006	6,729 (23)	36.5 (2,454)	**0.74 (0.67, 0.81)**
2007	2,878 (10)	32.6 (938)	**0.65 (0.58, 0.72)**
Age group at HIV diagnosis			
15-25	3,362 (11)	17.6 (593)	**0.32 (0.28, 0.35)**
25-29	5,798 (20)	27.5 (1,594)	**0.53 (0.49, 0.57)**
30-34	6,959 (23)	36.8 (2,559)	**0.81 (0.75, 0.87)**
35-39	5,780 (19)	41.4 (2,392)	-
40-44	3,435 (12)	46.1 (1,583)	**1.25 (1.15, 1.37)**
45-49	1,975 (7)	50.3 (994)	**1.51 (1.36, 1.68)**
50-54	1,147 (4)	51.1 (586)	**1.60 (1.40, 1.82)**
>54	1,239 (4)	54.5 (675)	**1.87 (1.65, 2.13)**
Region of diagnosis			
Outside London	13,827 (47)	37.4 (5,177)	**1.07 (1.02, 1.13)**
London	15,868 (53)	36.5 (5,799)	-
Risk group			
MSM	11,142 (38)	28.0 (3,121)	-
Heterosexual men	6,487 (22)	48.9 (3,174)	**1.77 (1.62, 1.92)**
Heterosexual women	9,333 (31)	42.3 (3,952)	**1.57 (1.44, 1.71)**
Heterosexual women (diagnosed antenatally)	2,034 (7)	22.0 (447)	**0.80 (0.70, 0.91)**
IDU	562 (2)	35.8 (201)	**1.48 (1.23, 1.77)**
Recipients of blood products	137 (0.5)	59.1 (81)	**2.54 (1.78, 3.63)**
Ethnicity			
White	12,734 (43)	30.2 (3,847)	-
Black African	13,319 (45)	43.4 (5,787)	**1.63 (1.50, 1.77)**
Black Caribbean	1,115 (4)	36.1 (402)	**1.23 (1.07, 1.41)**
Black Other	614 (2)	39.7 (244)	**1.55 (1.30, 1.86)**
Indian\Pakistani\Bangladeshi	431 (1)	38.5 (166)	1.23 (0.99, 1.51)
Other Asian	462 (2)	43.1 (199)	**2.06 (1.69, 2.51)**
Other/mixed	1,020 (3)	32.5 (331)	**1.22 (1.05, 1.41)**
Total	29,695 (100)	37.0 (10,976)	-

6.4.3 Sensitivity analyses for late diagnosis

6.4.3.1 Analysis of various definitions of late diagnosis

Increasing the CD4 count cut-off for the definition of late diagnosis to 350 cells/mm^3 increased the proportion considered to be diagnosed late to 60.0% (17,819/26,695) (Table 6.2). However, there was little change in the adjusted odds ratios for late diagnosis between groups in comparison to the previous analysis. The only marked difference between the analyses was that heterosexual women diagnosed antenatally were not significantly less likely to be diagnosed late than MSM using the cut-off of 350 cells/mm^3.

Defining late diagnosis as either AIDS within three months of an HIV diagnosis or CD4 counts less than 200 cells/mm^3 within 91 days increased the proportion considered to be diagnosed late to 39.5% (11,715/26,695) (Table 6.2). There were 737 individuals diagnosed with AIDS at CD4 counts greater than 199 cells/mm^3, with a median CD4 count of 300 (IQR 234, 418) cells/mm^3. There were no marked differences between the adjusted odds ratios between groups in this analysis compared to that without the inclusion of AIDS diagnoses.

Fewer than two-thirds (63.4% [18,835/29,695]) of individuals with their first CD4 cell counts within 91 days of HIV diagnosis had a second measurement within 91 days of the first (Table 6.3). The percentage of individuals considered to be diagnosed late increased from 32.9% (6,206), to 39.0% (7,338) and to 45.2% (8,521) as the definition was changed from both, the mean of, or either of, these measurements being less than 200 cells/mm^3. Although the percentage diagnosed late varied as the definition changed, the adjusted odds ratios of the

determining factors varied relatively little. However, definition of late diagnosis as both or the mean of the first two CD4 cell count less than 200 cells/mm^3 reduced the effects of the region and Indian\Pakistani\Bangladeshi ethnicity such that they were no longer statistically significant (Table 6.3).

Table 6.2. Proportion diagnosed late and odds ratios for late diagnosis using different definitions of late HIV diagnosis – individuals with CD4 cell counts within 91 days of their HIV diagnosis

	Late diagnosis defined as CD4 less than 350 cells/mm^3		Late diagnosis defined as CD4 less than 200 cells/mm^3 or AIDS diagnosis	
	Percentage of patients diagnosed late (CD4<200 cells per mm^3) % (n)	Multivariable adjusted odds ratio for late diagnosis (95% confidence interval)	Percentage of patients diagnosed late (CD4<200 cells per mm^3) % (n)	Multivariable adjusted odds ratio for late diagnosis (95% confidence interval)
Period of HIV diagnosis				
1995-1996	59.1 (1,006)	1.13 (1.00, 1.28)	41.2 (701)	**1.17 (1.03, 1.33)**
1997-1998	60.5 (1,339)	1.06 (0.94, 1.19)	43.2 (956)	1.11 (0.99, 1.25)
1999-2000	61.9 (1,983)	-	43 (1,377)	-
2001-2002	61.1 (3,341)	**0.86 (0.78, 0.94)**	40.8 (2,231)	**0.83 (0.75, 0.91)**
2003-2004	60 (4,505)	**0.79 (0.72, 0.86)**	38.2 (2,869)	**0.71 (0.65, 0.78)**
2005-2006	60.1 (4,046)	**0.79 (0.72, 0.87)**	38.5 (2,592)	**0.71 (0.65, 0.78)**
2007	55.6 (1,599)	**0.71 (0.64, 0.79)**	34.4 (989)	**0.62 (0.56, 0.70)**
Age group at HIV diagnosis				
15-25	42.4 (1,426)	**0.43 (0.39, 0.47)**	19.3 (649)	**0.32 (0.29, 0.36)**
25-29	52.4 (3,040)	**0.61 (0.56, 0.65)**	29.7 (1,723)	**0.54 (0.50, 0.58)**
30-34	61.2 (4,260)	**0.88 (0.82, 0.95)**	39.3 (2,734)	**0.82 (0.76, 0.88)**
35-39	63.4 (3,664)	-	43.8 (2,529)	-
40-44	67 (2,302)	**1.23 (1.13, 1.35)**	49.2 (1,689)	**1.30 (1.19, 1.41)**
45-49	71.6 (1,414)	**1.58 (1.41, 1.77)**	53.4 (1,055)	**1.56 (1.41, 1.74)**
50-54	69.6 (798)	**1.49 (1.29, 1.71)**	54 (619)	**1.64 (1.44, 1.87)**
>54	73.8 (915)	**1.90 (1.65, 2.19)**	57.9 (717)	**1.98 (1.74, 2.25)**
Region of diagnosis				
Outside London	60.7 (8,392)	**1.07 (1.02, 1.13)**	39.9 (5,516)	**1.07 (1.02, 1.13)**
London	59.4 (9,427)	-	39.1 (6,199)	-
Risk group				
MSM	48.9 (5,444)	-	29.9 (3,332)	-
Heterosexual men	71.2 (4,621)	**1.58 (1.46, 1.72)**	52.1 (3,379)	**1.77 (1.63, 1.92)**
Heterosexual women	66.8 (6,238)	**1.42 (1.31, 1.54)**	45.5 (4,242)	**1.57 (1.44, 1.71)**
Heterosexual women (diagnosed antenatally)	53 (1,078)	0.98 (0.88, 1.11)	22.7 (461)	**0.73 (0.64, 0.83)**
IDU	58.7 (330)	**1.51 (1.27, 1.80)**	38.1 (214)	**1.49 (1.25, 1.78)**
Recipients of blood products	78.8 (108)	**2.34 (1.53, 3.57)**	63.5 (87)	**2.71 (1.88, 3.90)**
Ethnicity				
White	50.5 (6,434)	-	32.1 (4,086)	-
Black African	69.5 (9,255)	**2.09 (1.93, 2.26)**	46.7 (6,215)	**1.74 (1.61, 1.89)**
Black Caribbean	57.3 (639)	**1.27 (1.11, 1.45)**	37.8 (421)	**1.22 (1.06, 1.41)**
Black Other	62.9 (386)	**1.70 (1.42, 2.03)**	41 (252)	**1.52 (1.28, 1.82)**
Indian\Pakistani\Bangladeshi	59.6 (257)	**1.29 (1.05, 1.59)**	41.5 (179)	**1.29 (1.05, 1.59)**
Other Asian	67.1 (310)	**2.27 (1.85, 2.79)**	45.2 (209)	**2.10 (1.72, 2.56)**
Other/mixed	52.7 (538)	**1.19 (1.04, 1.36)**	34.6 (353)	**1.24 (1.08, 1.43)**
Total	60.0 (17,819)	-	39.5 (11,715)	-

Table 6.3. Proportion diagnosed late and odds ratios for late diagnosis considering two CD4 cell counts in the definition of late HIV diagnosis – first within 91 days of their HIV diagnosis and second within 91 days of the first

	Percentage of individuals with CD4 within 91 days of HIV diagnosis included in this analysis	Multivariable adjusted odds ratio for late diagnosis (95% confidence interval)		
	% (n)	Late diagnosis defined as both first two CD4 counts less than 200 cells/mm^3	Late diagnosis defined as mean of first two CD4 counts less than 200 cells/mm^3	Late diagnosis defined as either of first two CD4 counts less than 200 cells/mm^3
Period of HIV diagnosis				
1995-1996	54.6	**1.31 (1.11, 1.55)**	**1.23 (1.05, 1.45)**	**1.25 (1.06, 1.47)**
1997-1998	57.7	1.11 (0.95, 1.29)	1.10 (0.95, 1.28)	1.12 (0.97, 1.30)
1999-2000	65.4	-	-	-
2001-2002	61.8	**0.84 (0.75, 0.95)**	**0.83 (0.74, 0.93)**	**0.85 (0.76, 0.96)**
2003-2004	66.3	**0.77 (0.68, 0.86)**	**0.77 (0.69, 0.86)**	**0.76 (0.69, 0.85)**
2005-2006	65.7	**0.78 (0.69, 0.87)**	**0.80 (0.71, 0.89)**	**0.83 (0.74, 0.93)**
2007	61.1	**0.69 (0.59, 0.79)**	**0.73 (0.64, 0.84)**	**0.74 (0.64, 0.84)**
Age group at HIV diagnosis				
15-25	57.5	**0.39 (0.34, 0.45)**	**0.36 (0.31, 0.41)**	**0.37 (0.32, 0.42)**
25-29	61.0	**0.62 (0.56, 0.69)**	**0.58 (0.53, 0.65)**	**0.57 (0.51, 0.63)**
30-34	63.6	**0.87 (0.79, 0.96)**	**0.84 (0.77, 0.92)**	**0.83 (0.76, 0.91)**
35-39	65.3	-	-	-
40-44	66.2	**1.23 (1.10, 1.38)**	**1.21 (1.09, 1.35)**	**1.27 (1.14, 1.41)**
45-49	66.5	**1.32 (1.16, 1.50)**	**1.41 (1.24, 1.60)**	**1.46 (1.28, 1.66)**
50-54	66.2	**1.59 (1.35, 1.87)**	**1.60 (1.36, 1.88)**	**1.56 (1.33, 1.84)**
>54	65.9	**1.57 (1.34, 1.84)**	**1.62 (1.39, 1.90)**	**1.73 (1.48, 2.03)**
Region of diagnosis				
Outside London	63.3	1.02 (0.95, 1.08)	1.06 (1.00, 1.13)	**1.18 (1.11, 1.26)**
London	63.5	-	-	-
Risk group				
MSM	64.8	-	-	-
Heterosexual men	61.8	**1.84 (1.65, 2.05)**	**1.88 (1.69, 2.08)**	**1.84 (1.65, 2.04)**
Heterosexual women	62.1	**1.55 (1.39, 1.73)**	**1.61 (1.45, 1.79)**	**1.62 (1.46, 1.79)**
Heterosexual women (diagnosed antenatally)	70.4	**0.55 (0.46, 0.66)**	**0.57 (0.49, 0.68)**	**0.64 (0.55, 0.75)**
IDU	50.0	**1.59 (1.23, 2.05)**	**1.61 (1.26, 2.07)**	**1.69 (1.33, 2.16)**
Recipients of blood products	65.7	**2.40 (1.56, 3.69)**	**2.53 (1.63, 3.93)**	**2.42 (1.54, 3.81)**
Ethnicity				
White	64.3	-	-	-
Black African	63.0	**1.63 (1.47, 1.81)**	**1.59 (1.44, 1.76)**	**1.65 (1.49, 1.82)**
Black Caribbean	58.2	**1.42 (1.18, 1.70)**	**1.31 (1.09, 1.57)**	**1.25 (1.05, 1.49)**
Black Other	61.6	**1.66 (1.32, 2.09)**	**1.61 (1.28, 2.01)**	**1.61 (1.29, 2.02)**
Indian\Pakistani\Bangladeshi	61.3	1.12 (0.85, 1.48)	1.30 (1.00, 1.69)	**1.42 (1.09, 1.85)**
Other Asian	68.2	**2.29 (1.80, 2.91)**	**1.99 (1.56, 2.52)**	**2.04 (1.61, 2.59)**
Other/mixed	63.8	**1.27 (1.06, 1.53)**	**1.23 (1.03, 1.47)**	**1.30 (1.10, 1.55)**
Total	63.4	-	-	-

6.4.3.2 Analysis of various definitions of late diagnosis using different cut-off times from HIV diagnosis to first CD4 cell count

Increasing the time between HIV diagnosis and first CD4 cell count used for the definition of late diagnosis increased the percentage and number of individuals that were included in the analysis (Section 5.4.4.1). For cut-offs of 30 days, 60 days and 182 days between first CD4 cell count and HIV diagnosis, the percentages and numbers included were 71.7% (24,869), 81.9% (28,405) and 90.5% (31,390) of all 34,692 individuals in the integrated dataset who were diagnosed between 1995 and 2007 (Table 6.4). However, the different cut-offs had little effect on the adjusted odds ratios for late diagnosis.

Table 6.4. Proportion diagnosed late and odds ratios for late diagnosis defined by CD4 counts less than 200 cells/mm^3 within 30 days, 60 days and 182 days of their HIV diagnosis

	Multivariable adjusted odds ratio for late diagnosis (95% confidence interval)		
	Late diagnosis defined as CD4 < 200 cells/mm^3 within 30 days of HIV diagnosis	**Late diagnosis defined as CD4 < 200 cells/mm^3 within 60 days of HIV diagnosis**	**Late diagnosis defined as CD4 < 200 cells/mm^3 within 182 days of HIV diagnosis**
Percentage (number) of all individuals in the analysis	**71.7% (24,869)**	**81.9% (28,405)**	**90.5% (31,390)**
Period of HIV diagnosis			
1995-1996	1.07 (0.93, 1.23)	1.11 (0.97, 1.26)	**1.15 (1.02, 1.29)**
1997-1998	1.03 (0.90, 1.17)	1.06 (0.94, 1.20)	1.09 (0.97, 1.21)
1999-2000	-	-	-
2001-2002	**0.83 (0.75, 0.92)**	**0.82 (0.74, 0.90)**	**0.83 (0.76, 0.91)**
2003-2004	**0.74 (0.67, 0.81)**	**0.73 (0.66, 0.80)**	**0.73 (0.67, 0.80)**
2005-2006	**0.76 (0.69, 0.84)**	**0.73 (0.66, 0.80)**	**0.74 (0.68, 0.81)**
2007	**0.66 (0.58, 0.74)**	**0.65 (0.58, 0.72)**	**0.65 (0.58, 0.72)**
Age group at HIV diagnosis			
15-25	**0.31 (0.28, 0.35)**	**0.31 (0.28, 0.35)**	**0.33 (0.29, 0.36)**
25-29	**0.53 (0.48, 0.58)**	**0.53 (0.49, 0.58)**	**0.54 (0.50, 0.58)**
30-34	**0.82 (0.76, 0.89)**	**0.82 (0.76, 0.88)**	**0.82 (0.76, 0.88)**
35-39	-	-	-
40-44	**1.26 (1.15, 1.39)**	**1.25 (1.15, 1.37)**	**1.26 (1.15, 1.37)**
45-49	**1.50 (1.34, 1.68)**	**1.51 (1.36, 1.69)**	**1.50 (1.36, 1.67)**
50-54	**1.57 (1.36, 1.80)**	**1.58 (1.38, 1.80)**	**1.62 (1.43, 1.84)**
>54	**1.90 (1.65, 2.18)**	**1.89 (1.66, 2.15)**	**1.88 (1.66, 2.13)**
Region of diagnosis			
Outside London	**1.06 (1.00, 1.12)**	**1.07 (1.01, 1.13)**	**1.08 (1.03, 1.14)**
London	-	-	-
Risk group			
MSM	-	-	-
Heterosexual men	**1.74 (1.58, 1.90)**	**1.76 (1.62, 1.92)**	**1.80 (1.66, 1.95)**
Heterosexual women	**1.58 (1.44, 1.73)**	**1.58 (1.45, 1.73)**	**1.58 (1.45, 1.72)**
Heterosexual women (diagnosed antenatally)	**0.82 (0.71, 0.96)**	**0.81 (0.71, 0.93)**	**0.82 (0.72, 0.94)**
IDU	**1.56 (1.27, 1.91)**	**1.52 (1.26, 1.84)**	**1.49 (1.25, 1.77)**
Recipients of blood products	**2.32 (1.56, 3.44)**	**2.41 (1.68, 3.47)**	**2.80 (1.98, 3.97)**
Ethnicity			
White	-	-	-
Black African	**1.61 (1.47, 1.76)**	**1.62 (1.49, 1.76)**	**1.62 (1.50, 1.76)**
Black Caribbean	**1.20 (1.03, 1.40)**	**1.24 (1.07, 1.43)**	**1.23 (1.07, 1.40)**
Black Other	**1.46 (1.20, 1.78)**	**1.51 (1.26, 1.81)**	**1.53 (1.28, 1.82)**
Indian\Pakistani\Bangladeshi	1.16 (0.92, 1.46)	1.23 (1.00, 1.53)	1.19 (0.96, 1.46)
Other Asian	**2.04 (1.65, 2.53)**	**2.05 (1.68, 2.51)**	**2.02 (1.66, 2.46)**
Other/mixed	1.15 (0.99, 1.35)	**1.21 (1.04, 1.40)**	**1.22 (1.06, 1.41)**

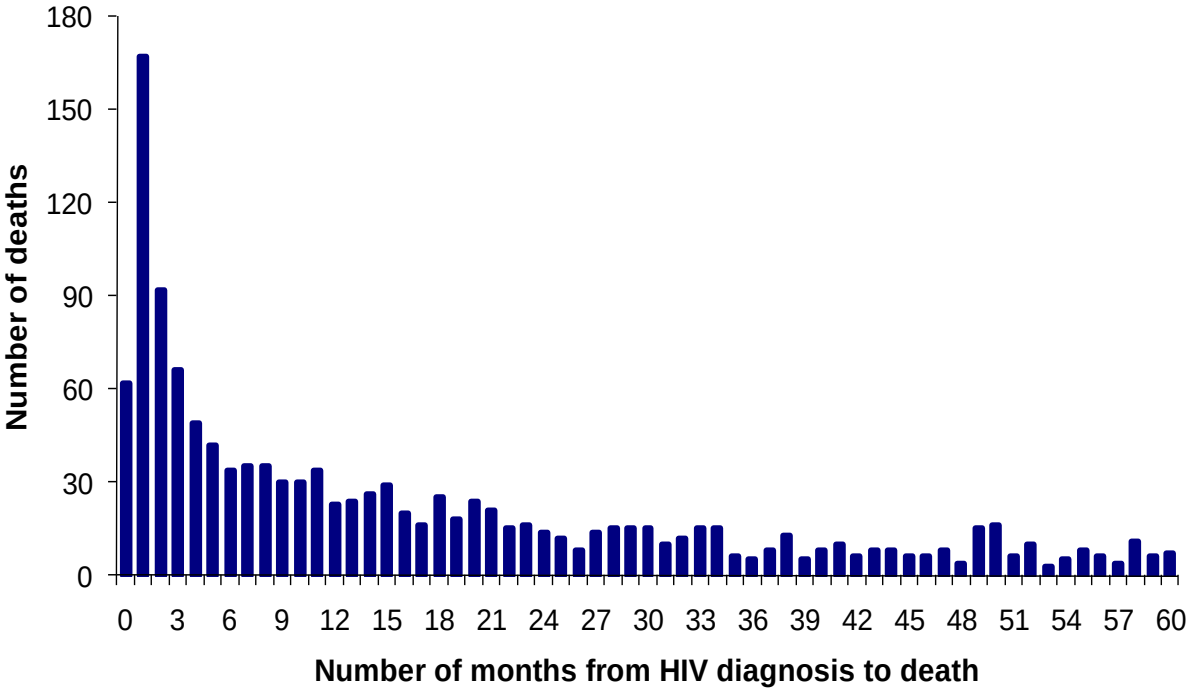
6.4.4 Short-term mortality

Estimated short-term mortality accounted for 17.9% (744/4,183) of all deaths occurring among this population between 1995 and 2007. Although short-term mortality declined when HAART became available, it accounted for an increasing percentage of all deaths between 1995 and 2001 (7.9% (47/597) in 1995 to 28.5% (68/239) in 2001) before stabilising at an average of 26.4% between 2001 and 2007. Short-term mortality between 2000 and 2007 was 1.9% (505) of all 27,201 newly diagnosed individuals and 2.0% (483/24,444) of those with CD4 cell counts within 91 days of HIV diagnosis.

The number of deaths among all individuals was particularly high during the first three months after HIV diagnosis and then declined gradually but remained elevated for the first two years after HIV diagnosis in comparison to the low rate thereafter (Figure 6.4). Overall short-term mortality among all individuals diagnosed between 1995 and 2007 was 2.08% (721/34,692) (Table 6.5). The mortality percentages within one, three, six, 12 and 24 months of diagnosis among all individuals diagnosed between 1995 and 2007 were 0.35% (123), 1.01% (350), 1.46% (505), 2.08% (721) and 2.93% (1,018) respectively.

The factors associated with short-term mortality were similar to those for late diagnosis. Short-term mortality declined over time from 3.6% (98/2,697) in 1995/6 to 1.1% (34/2,968) in 2007 (AOR 0.25, 95% CI 0.16, 0.37) with significant drops between 1995/6 and 1997/8 (AOR 0.61, 95% CI 0.45, 0.83) and between 2005/6 and 2007 (AOR 0.62, 95% CI 0.42, 0.90). The percentage of deaths

Figure 6.4. Number of deaths among all individuals who were known to have died by the time from HIV diagnosis to death (deaths within five years of HIV diagnosis and 12 months of a CD4 cell count)



within a year of HIV diagnosis increased markedly from 0.5% (20/4,037) among those aged 15-25 years to 6.9% (97/1,399) among those aged older than 54 years (AOR 13.35, 95% CI 8.17, 21.81). Short-term mortality was higher among individuals diagnosed outside London than in London (AOR 1.24, 95% CI 1.06, 1.45). Heterosexual men (AOR 1.38, 95% CI 1.09, 1.74) and IDU (AOR 2.30, 95% CI 1.53, 3.46) were more likely to die within a year of HIV diagnosis than MSM but heterosexual women diagnosed antenatally were least likely to do so (AOR 0.17, 95% CI 0.06, 0.47). Short-term mortality was more likely for Black African (AOR 1.27, 95% CI 1.019, 1.61) and Indian/Pakistani/Bangladeshi (AOR 1.71, 95% CI 1.04, 2.80) individuals.

Table 6.5. Short-term mortality and odds ratios for short-term mortality among all individuals diagnosed between 1995 and 2007

	Short-term mortality among all patients (total number diagnosed) **% (n)**	**Multivariable adjusted odds ratio for late diagnosis (95% confidence interval)**
Period of HIV diagnosis		
1995-1996	3.6 (98 / 2,697)	**1.57 (1.18, 2.09)**
1997-1998	2.5 (77 / 3,062)	0.95 (0.71, 1.29)
1999-2000	2.6 (103 / 3,923)	-
2001-2002	1.9 (121 / 6,418)	**0.70 (0.53, 0.91)**
2003-2004	1.9 (156 / 8,280)	**0.69 (0.53, 0.89)**
2005-2006	1.8 (132 / 7,344)	**0.62 (0.48, 0.81)**
2007	1.1 (34 / 2,968)	**0.38 (0.26, 0.57)**
Age group at HIV diagnosis		
15-25	0.5 (20 / 4,037)	**0.26 (0.16, 0.42)**
25-29	1.0 (70 / 6,983)	**0.50 (0.37, 0.66)**
30-34	1.6 (130 / 8,137)	**0.76 (0.60, 0.97)**
35-39	2.1 (139 / 6,630)	-
40-44	2.8 (109 / 3,933)	**1.35 (1.05, 1.74)**
45-49	4.0 (91 / 2,255)	**1.99 (1.52, 2.60)**
50-54	4.9 (65 / 1,318)	**2.40 (1.77, 3.26)**
>54	6.9 (97 / 1,399)	**3.50 (2.67, 4.60)**
Region of diagnosis		
Outside London	2.3 (360 / 15,935)	**1.24 (1.06, 1.45)**
London	1.9 (361 / 18,757)	-
Risk group		
MSM	1.9 (248 / 13,124)	-
Heterosexual men	3.1 (233 / 7,584)	**1.38 (1.09, 1.74)**
Heterosexual women	1.8 (202 / 10,920)	1.05 (0.81, 1.36)
Heterosexual women (diagnosed antenatally)	0.2 (4 / 2,187)	**0.17 (0.06, 0.47)**
IDU	3.8 (27 / 713)	**2.30 (1.53, 3.46)**
Recipients of blood products	4.3 (7 / 164)	1.49 (0.68, 3.30)
Ethnicity		
White	2.2 (331 / 15,092)	-
Black African	2.0 (308 / 15,423)	**1.27 (1.01, 1.61)**
Black Caribbean	1.8 (23 / 1,312)	0.97 (0.62, 1.51)
Black Other	2.0 (14 / 693)	1.37 (0.78, 2.40)
Indian\Pakistani\Bangladeshi	3.8 (19 / 499)	**1.71 (1.04, 2.80)**
Other Asian	2.2 (11 / 506)	1.49 (0.80, 2.78)
Other/mixed	1.3 (15 / 1,167)	0.80 (0.47, 1.36)
Total	2.1 (721 / 34,692)	-

The majority (86.0% [590/686]) of those who died within a year of diagnosis were diagnosed late. The number of deaths among individuals diagnosed late was particularly high during the first three months after HIV diagnosis but remained elevated for the first two years after HIV diagnosis in comparison to deaths among those diagnosed with higher CD4 cell counts (Figure 6.5). Among individuals diagnosed late, the percentage who had died increased from 1.06% (116) within one month of diagnosis to 7.07 (776) within two years of diagnosis. In contrast, these percentages were 0.04% (7) and 0.88% (165) among individuals diagnosed with higher CD4 cell counts (Table 6.6).

Figure 6.5. Number of deaths among those diagnosed late and those diagnosed earlier by time from HIV diagnosis to death (deaths within three years of HIV diagnosis)

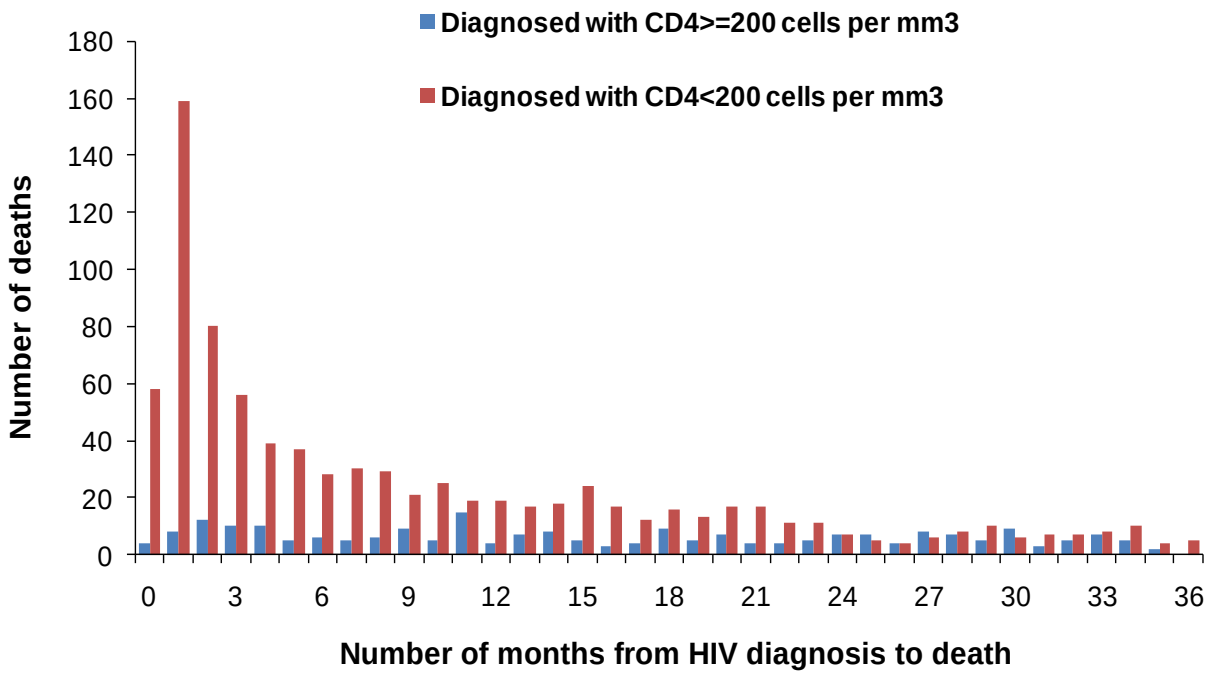
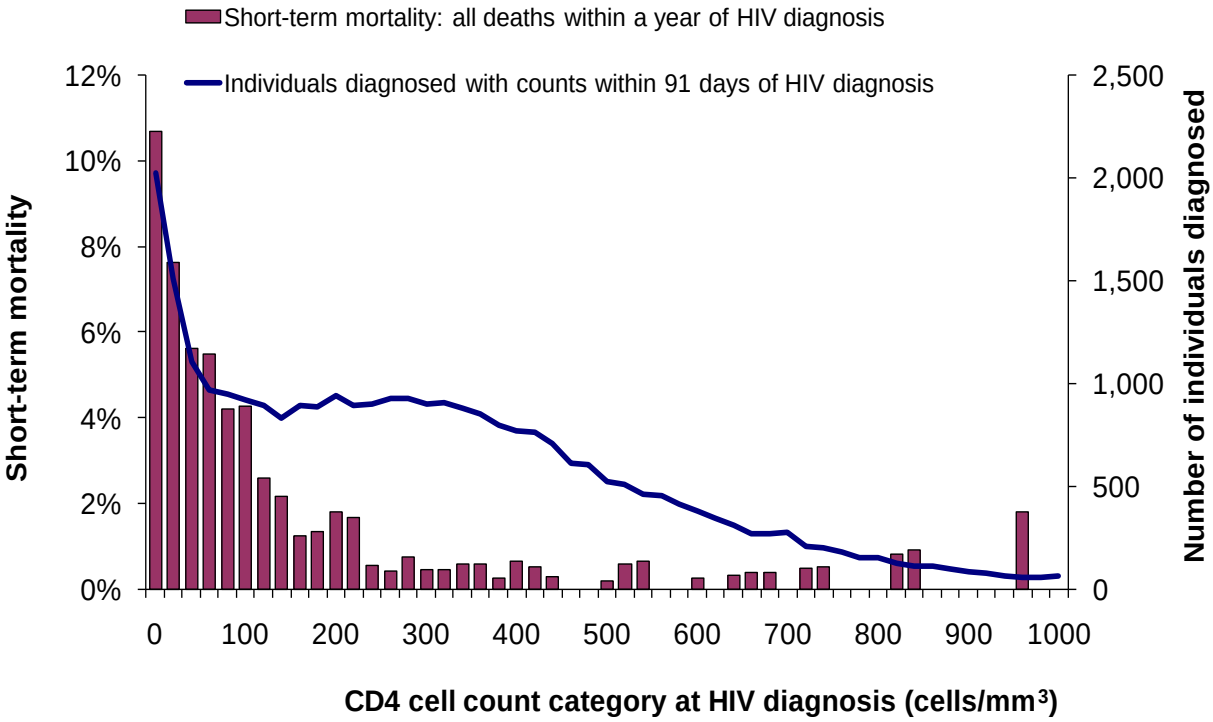


Table 6.6. Percentage of deaths among those diagnosed late and those diagnosed earlier by time from HIV diagnosis to death

	Percentage of deaths among individuals newly diagnosed	
Time from HIV diagnosis (months)	**Diagnosed with CD4 < 200 cells/mm^3**	**Diagnosed with CD4 ≥ 200 cells/mm^3**
1	1.06 (116)	0.04 (7)
3	2.92 (320)	0.16 (30)
6	4.05 (444)	0.27 (51)
12	5.38 (590)	0.51 (96)
24	7.07 (776)	0.88 (165)

Short-term mortality (i.e. deaths within the first year) decreased with increasing CD4 cell counts at the time of HIV diagnosis for those individuals with measurements within 91 days of their HIV diagnosis (Figure 6.6). Short-term mortality was relatively high for individuals diagnosed with CD4 counts between 0 and 239 cells/mm^3 but decreased from 10.7% at CD4 counts between 0 and 19 cells/mm^3 to 1.7% at CD4 counts between 220 and 239 cells/mm^3.

Figure 6.6. Short-term mortality and numbers diagnosed by CD4 cell count categories at HIV diagnosis (CD4 cell counts grouped from 0-19 cells/mm^3, 20-39 cells/mm^3 etc.)



Individuals diagnosed late were eleven times more likely to die from any cause within a year of HIV diagnosis than individuals diagnosed more promptly (5.4% [590/10,976] vs 0.5% [96/18,719]) (Table 6.7). This difference was marked among all groups. The independent effect of late diagnosis on short-term mortality, after controlling for other factors, remained strong (AOR 8.71, 95% CI 6.97, 10.87). Period and region of HIV diagnosis and age group were significantly associated with short-term mortality in multivariable analysis. Heterosexual men did not have significantly higher odds of short-term mortality than MSM although there was a weak residual association (AOR 1.26, 95% CI 0.99, 1.61). Heterosexual women were less likely to die within a year of HIV

diagnosis than MSM (OR 0.21, 95% CI 0.08, 0.59) but IDU had slightly higher odds of short-term mortality compared to MSM even after controlling for late diagnosis (AOR 2.39, 95% CI 1.56, 3.67). There were no statistically significant differences in the odds of short-term mortality between white individuals and individuals in other ethnic groups suggesting that differences in short-term mortality were explained by different proportions diagnosed late.

Table 6.7. Short-term mortality and factors associated with short-term mortality among individuals with CD4 cell counts within 91 days of their HIV diagnosis

	Short-term mortality among patients with CD4 cell counts at HIV diagnosis		Multivariable adjusted odds ratio for short-term mortality
	CD4 < 200 % (n)	CD4 ≥ 200 % (n)	(95% confidence interval)
Late HIV diagnosis status			
CD4 < 200	5.4 (590)	-	**8.71 (6.97, 10.87)**
CD4 ≥ 200	-	0.5 (96)	-
Period of HIV diagnosis			
1995-1996	12.6 (82)	1.0 (10)	**2.13 (1.56, 2.90)**
1997-1998	7.7 (68)	0.4 (5)	1.12 (0.81, 1.55)
1999-2000	5.7 (73)	1.0 (19)	-
2001-2002	5.0 (104)	0.4 (14)	0.78 (0.59, 1.04)
2003-2004	4.5 (121)	0.6 (28)	0.76 (0.58, 1.00)
2005-2006	4.6 (113)	0.4 (15)	**0.69 (0.52, 0.91)**
2007	3.1 (29)	0.3 (5)	**0.44 (0.29, 0.66)**
Age group at HIV diagnosis			
15-25	2.5 (15)	0.1 (4)	**0.49 (0.30, 0.80)**
25-29	3.4 (54)	0.3 (12)	**0.71 (0.52, 0.96)**
30-34	3.9 (100)	0.5 (22)	0.85 (0.66, 1.09)
35-39	4.5 (108)	0.7 (23)	-
40-44	5.9 (93)	0.6 (11)	1.25 (0.96, 1.63)
45-49	7.9 (79)	1.0 (10)	**1.76 (1.33, 2.33)**
50-54	9.2 (54)	1.2 (7)	**2.01 (1.46, 2.78)**
>54	12.9 (87)	1.2 (7)	**2.85 (2.14, 3.80)**
Region of diagnosis			
Outside London	5.8 (298)	0.5 (45)	**1.18 (1.00, 1.38)**
London	5.0 (292)	0.5 (51)	-
Risk group			
MSM	6.2 (194)	0.5 (38)	-
Heterosexual men	6.2 (198)	0.8 (27)	1.26 (0.99, 1.61)
Heterosexual women	4.3 (170)	0.4 (22)	0.99 (0.75, 1.29)
Heterosexual women (diagnosed antenatally)	0.4 (2)	0.1 (2)	**0.21 (0.08, 0.59)**
IDU	10.4 (21)	1.7 (6)	**2.39 (1.56, 3.67)**
Recipients of blood products	6.2 (5)	1.8 (1)	1.01 (0.42, 2.39)
Ethnicity			
White	6.9 (266)	0.6 (49)	-
Black African	4.4 (256)	0.5 (39)	1.00 (0.78, 1.28)
Black Caribbean	4.7 (19)	0 (0)	0.73 (0.45, 1.19)
Black Other	4.1 (10)	0.8 (3)	1.04 (0.58, 1.88)
Indian\Pakistani\Bangladeshi	10.8 (18)	0.4 (1)	1.57 (0.94, 2.63)
Other Asian	4.5 (9)	0.8 (2)	1.13 (0.60, 2.13)
Other/mixed	3.6 (12)	0.3 (2)	0.67 (0.39, 1.18)

6.4.5 Sensitivity analyses for short-term mortality

6.4.5.1 Analysis of various definitions of short-term mortality using different cut-off times from HIV diagnosis to death

Varying the cut-off time between HIV diagnosis and death for the definition of short-term mortality significantly changed some of the comparative effects of the associated factors (Table 6.8). Increasing the time between diagnosis and death from three months to 24 months resulted in a decrease in the adjusted odds ratio for late diagnosis from 15.41 (95% CI 10.52, 22.56) to 6.99 (95% CI 5.87, 8.33). In contrast, the effects of period of HIV diagnosis, age group and risk group became more significant as the cut-off increased. Compared to 1999/2000, there was no significant difference in the odds of death within three months of diagnosis for any period of HIV diagnosis except 1995/6, when death was more likely (AOR 1.72; 95% CI 1.07, 2.74). For deaths within six months of diagnosis, diagnosis in 2007 was associated with significantly lower odds of death than diagnosis in 1999/2000 (AOR 0.55; 95% CI 0.35, 0.86). For deaths within 24 months of diagnosis, there were also significant differences for diagnoses in 1997/8 and 2005/6 compared to 1999/2000 and the odds ratios for diagnoses 1995/6 and 2007 were greater. Being aged 45 years or older was associated with significantly higher odds of death after diagnosis irrespective of the cut-off used. However, the odds of death when diagnosed aged 40-44 years compared to those when aged 35-39 years became significantly different as the cut-off increased from three months to six months and there was evidence of significantly lower odds of death among younger individuals as the cut-off increased to more than six months. Interestingly, mortality within three months

of diagnosis was not significantly different for MSM than for any other risk group. Consideration of deaths within 24 months of diagnosis indicated that heterosexual men had significantly higher odds of mortality than MSM (AOR 1.24; 95% CI 1.00, 1.54). Diagnosis outside London was consistently associated with higher mortality whereas there were consistently no significant differences in mortality between white individuals and other ethnic groups.

Table 6.8. Factors associated with short-term mortality among individuals with CD4 cell counts within 91 days of their HIV diagnosis: considering different cut-offs of time from HIV diagnosis to death

	Multivariable adjusted odds ratio for short-term mortality (95% confidence interval)		
	Deaths within three months of diagnosis	**Deaths within six months of diagnosis**	**Deaths within 24 months of diagnosis**
Late HIV diagnosis status			
CD4 < 200	**15.41 (10.52, 22.56)**	**12.63 (9.39, 16.99)**	**6.99 (5.87, 8.33)**
CD4 ≥ 200	-	-	-
Period of HIV diagnosis			
1995-1996	**1.72 (1.07, 2.74)**	**1.59 (1.09, 2.32)**	**3.24 (2.50, 4.21)**
1997-1998	1.39 (0.89, 2.17)	1.13 (0.78, 1.64)	**1.40 (1.07, 1.85)**
1999-2000	-	-	-
2001-2002	0.79 (0.52, 1.20)	0.77 (0.55, 1.08)	0.84 (0.65, 1.08)
2003-2004	0.97 (0.66, 1.42)	0.81 (0.59, 1.11)	0.81 (0.64, 1.04)
2005-2006	0.98 (0.66, 1.44)	0.72 (0.52, 1.00)	**0.68 (0.53, 0.88)**
2007	0.64 (0.37, 1.10)	**0.55 (0.35, 0.86)**	**0.34 (0.23, 0.51)**
Age group at HIV diagnosis			
15-25	0.74 (0.39, 1.39)	0.72 (0.42, 1.22)	**0.51 (0.34, 0.77)**
25-29	0.87 (0.57, 1.33)	0.89 (0.63, 1.27)	0.82 (0.64, 1.06)
30-34	0.90 (0.63, 1.30)	0.89 (0.65, 1.21)	0.90 (0.72, 1.12)
35-39	-	-	-
40-44	1.26 (0.86, 1.84)	**1.37 (1.00, 1.88)**	**1.32 (1.05, 1.67)**
45-49	**2.11 (1.44, 3.09)**	**1.92 (1.38, 2.68)**	**1.92 (1.50, 2.46)**
50-54	**1.92 (1.21, 3.05)**	**1.96 (1.33, 2.89)**	**2.04 (1.54, 2.72)**
>54	**3.55 (2.42, 5.20)**	**3.29 (2.36, 4.57)**	**3.02 (2.34, 3.90)**
Region of diagnosis			
Outside London	**1.38 (1.11, 1.72)**	**1.24 (1.03, 1.50)**	**1.17 (1.02, 1.34)**
London	-	-	-
Risk group			
MSM	-	-	-
Heterosexual men	1.06 (0.76, 1.49)	1.25 (0.95, 1.66)	**1.24 (1.00, 1.54)**
Heterosexual women	0.87 (0.60, 1.26)	0.98 (0.71, 1.33)	0.98 (0.77, 1.23)
Heterosexual women (diagnosed antenatally)	0.37 (0.13, 1.06)	**0.29 (0.11, 0.82)**	**0.31 (0.15, 0.65)**
IDU	1.76 (0.93, 3.33)	**2.29 (1.39, 3.77)**	**2.82 (1.99, 4.00)**
Recipients of blood products	0.79 (0.24, 2.60)	1.13 (0.44, 2.91)	0.73 (0.31, 1.74)
Ethnicity			
White	-	-	-
Black African	1.04 (0.74, 1.45)	0.91 (0.68, 1.20)	0.98 (0.79, 1.21)
Black Caribbean	0.87 (0.45, 1.66)	0.71 (0.40, 1.26)	0.76 (0.50, 1.15)
Black Other	0.49 (0.15, 1.58)	1.06 (0.54, 2.07)	0.98 (0.59, 1.65)
Indian\Pakistani\Bangladeshi	1.65 (0.83, 3.29)	1.50 (0.83, 2.70)	1.56 (0.99, 2.46)
Other Asian	1.17 (0.50, 2.74)	1.17 (0.58, 2.36)	0.97 (0.54, 1.75)
Other/mixed	0.82 (0.39, 1.71)	0.73 (0.39, 1.36)	0.71 (0.45, 1.12)

6.4.5.2 Different categorisation of CD4 cell counts

Determination of the association between CD4 cell counts and short-term mortality by analysis of the effect of CD4 count categories showed that there was a non-linear decrease in the odds of death within a year of diagnosis with increasing CD4 count category (Table 6.9).

Neither the square root transformation or the base ten logarithm transformation of the CD4 cell count fully reflected the high short-term mortality at very low CD4 cell counts, the lower short-term mortality at CD4 counts over 100 cells/mm^3, and the continuing all-cause mortality at high CD4 counts (Figures 6.7 and 6.8). However, the log transformation more closely reflected the association with short-term mortality. Analysis showed that an increase of one log in CD4 counts at the time of HIV diagnosis was associated with a five-fold decrease in the odds of death within a year of HIV diagnosis.

Consideration of the effect of late diagnosis, as defined by CD4 counts less than 350 cells/mm^3, on short-term mortality indicated slightly higher odds than with the definition of CD4 counts less than 200 cells/mm^3 (AOR 9.65, 95% CI 6.90, 13.50 compared to AOR 8.71, 95% CI 6.97, 10.87). The odds ratio for short-term mortality was even greater when AIDS diagnoses were included in the definition of late diagnosis along with CD4 counts less than 200 cells/mm^3 (AOR 13.93, 95% CI 10.58, 18.33).

The different categorisation of CD4 cell counts or late diagnosis in the various multivariable models had little effect on the odds ratios for the other factors.

Individuals diagnosed outside London had marginally higher odds of short-term mortality than individuals diagnosed in London but this was only statistically significant when late diagnosis was included in the model defined by CD4 counts less than 350 cells/mm^3. A similar pattern was observed for the odds of short-term mortality for heterosexual men.

Table 6.9. Effect of CD4 cell counts and late diagnosis on short-term mortality among individuals with CD4 cell counts within 91 days of their HIV diagnosis

	Multivariable adjusted odds ratio for short-term mortality (95% confidence interval)			
	CD4 cell count categories	**Log CD4 at diagnosis**	**Late diagnosis defined by CD4 less than 350 cells/mm^3**	**Late diagnosis defined by CD4 less than 200 cells/mm^3 or AIDS at diagnosis**
CD4 category at diagnosis				
<50	**5.37 (3.79, 7.62)**			
50-99	**2.63 (1.79, 3.85)**			
100-149	**1.89 (1.25, 2.84)**			
150-199	0.90 (0.56, 1.46)			
200-249	-			
250-299	**0.36 (0.19, 0.70)**			
300-349	**0.34 (0.17, 0.67)**			
350-499	**0.26 (0.15, 0.45)**			
>499	**0.19 (0.11, 0.34)**			
Log CD4 at diagnosis		**0.22 (0.20, 0.24)**		
CD4 < 350 cells/mm^3 at diagnosis				
CD4 < 350			**9.65 (6.90, 13.50)**	
CD4 ≥ 350			-	
CD4 < 200 cells/mm^3 or AIDS at diagnosis				
CD4 < 200 or AIDS				**13.93 (10.58, 18.33)**
CD4 ≥ 200 and no AIDS				-

Period of HIV diagnosis				
1995-1996	**2.05 (1.50, 2.80)**	**2.11 (1.54, 2.90)**	**2.08 (1.53, 2.82)**	**2.08 (1.53, 2.84)**
1997-1998	1.06 (0.76, 1.46)	1.04 (0.75, 1.44)	1.13 (0.82, 1.55)	1.09 (0.79, 1.51)
1999-2000	-	-	-	-
2001-2002	0.79 (0.59, 1.05)	0.80 (0.60, 1.07)	0.76 (0.57, 1.01)	0.78 (0.59, 1.04)
2003-2004	0.78 (0.59, 1.03)	**0.75 (0.56, 0.99)**	**0.72 (0.55, 0.94)**	0.77 (0.59, 1.02)
2005-2006	**0.69 (0.52, 0.92)**	**0.64 (0.48, 0.85)**	**0.65 (0.49, 0.86)**	**0.70 (0.53, 0.93)**
2007	**0.45 (0.30, 0.68)**	**0.42 (0.28, 0.63)**	**0.41 (0.27, 0.61)**	**0.45 (0.30, 0.68)**
Age group at HIV diagnosis				
15-25	**0.54 (0.33, 0.89)**	**0.48 (0.29, 0.78)**	**0.40 (0.24, 0.65)**	**0.52 (0.32, 0.85)**
25-29	0.75 (0.55, 1.01)	**0.69 (0.51, 0.95)**	**0.63 (0.47, 0.85)**	**0.72 (0.53, 0.98)**
30-34	0.86 (0.67, 1.11)	0.83 (0.64, 1.07)	0.81 (0.63, 1.04)	0.85 (0.66, 1.10)
35-39	-	-	-	-
40-44	1.26 (0.96, 1.64)	1.31 (1.00, 1.71)	1.30 (1.00, 1.69)	1.23 (0.94, 1.60)
45-49	**1.75 (1.32, 2.33)**	**1.89 (1.42, 2.51)**	**1.81 (1.37, 2.40)**	**1.72 (1.30, 2.28)**
50-54	**1.97 (1.42, 2.72)**	**2.12 (1.53, 2.94)**	**2.17 (1.58, 2.98)**	**1.98 (1.44, 2.73)**
>54	**2.97 (2.23, 3.97)**	**3.28 (2.45, 4.39)**	**3.04 (2.29, 4.03)**	**2.75 (2.07, 3.66)**
Region of diagnosis				
Outside London	1.16 (0.99, 1.36)	1.16 (0.99, 1.37)	**1.20 (1.02, 1.41)**	1.17 (1.00, 1.37)
London	-	-	-	-
Risk group				
MSM	-	-	-	-
Heterosexual men	1.16 (0.91, 1.48)	1.17 (0.91, 1.50)	**1.38 (1.08, 1.76)**	1.24 (0.97, 1.59)
Heterosexual women	0.92 (0.70, 1.21)	0.91 (0.69, 1.20)	1.04 (0.80, 1.37)	0.98 (0.75, 1.28)
Heterosexual women (diagnosed antenatally)	**0.26 (0.09, 0.72)**	**0.25 (0.09, 0.70)**	**0.19 (0.07, 0.51)**	**0.23 (0.08, 0.63)**
IDU	**2.29 (1.48, 3.53)**	**2.26 (1.46, 3.51)**	**2.39 (1.57, 3.63)**	**2.37 (1.55, 3.64)**
Recipients of blood products	0.96 (0.40, 2.27)	1.00 (0.42, 2.42)	1.11 (0.47, 2.64)	0.95 (0.40, 2.26)
Ethnicity				
White	-	-	-	-
Black African	1.06 (0.83, 1.36)	1.06 (0.82, 1.36)	0.99 (0.78, 1.27)	0.96 (0.75, 1.22)
Black Caribbean	0.70 (0.43, 1.15)	0.71 (0.43, 1.16)	0.75 (0.46, 1.22)	0.73 (0.44, 1.19)
Black Other	1.00 (0.55, 1.80)	1.01 (0.56, 1.84)	1.07 (0.59, 1.92)	1.04 (0.58, 1.87)
Indian\Pakistani\Bangladeshi	1.61 (0.96, 2.71)	1.69 (1.00, 2.86)	1.57 (0.95, 2.61)	1.51 (0.90, 2.53)
Other Asian	0.99 (0.52, 1.87)	1.01 (0.53, 1.93)	1.19 (0.63, 2.23)	1.10 (0.58, 2.07)
Other/mixed	0.67 (0.38, 1.18)	0.71 (0.40, 1.24)	0.70 (0.40, 1.22)	0.66 (0.38, 1.15)

Figure 6.7. Short-term mortality and numbers diagnosed by logarithm (base 10) of the CD4 cell count at HIV diagnosis (log CD4 less or equal to three)

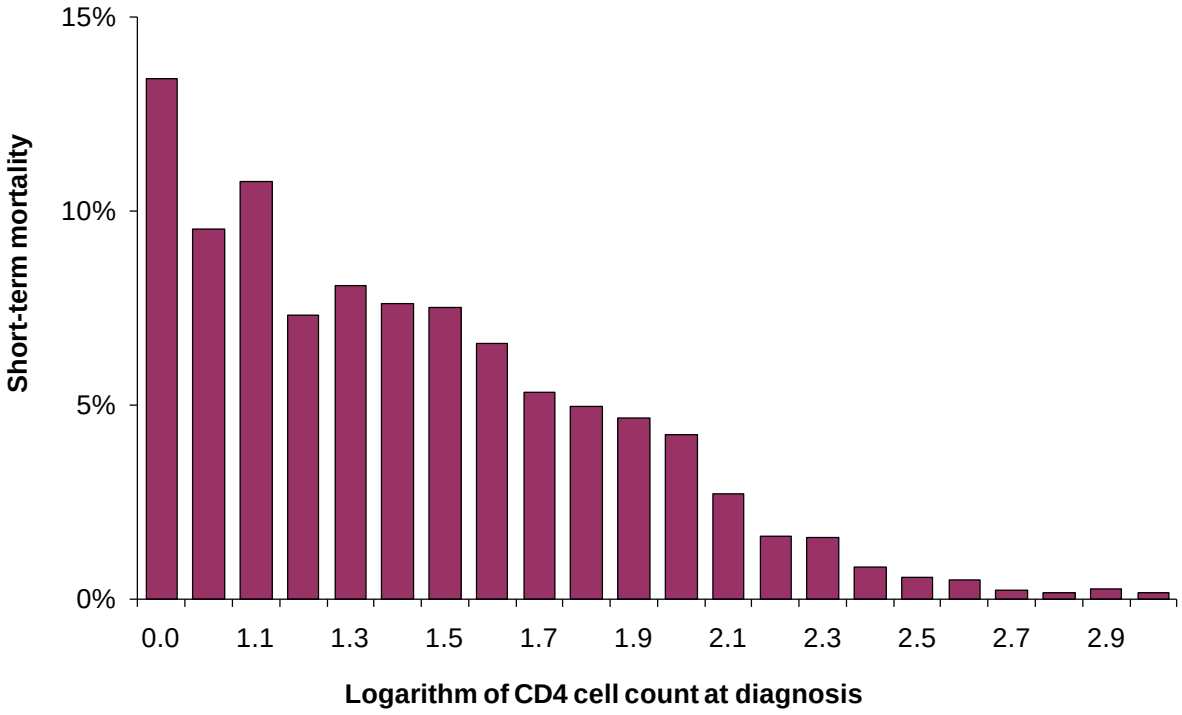
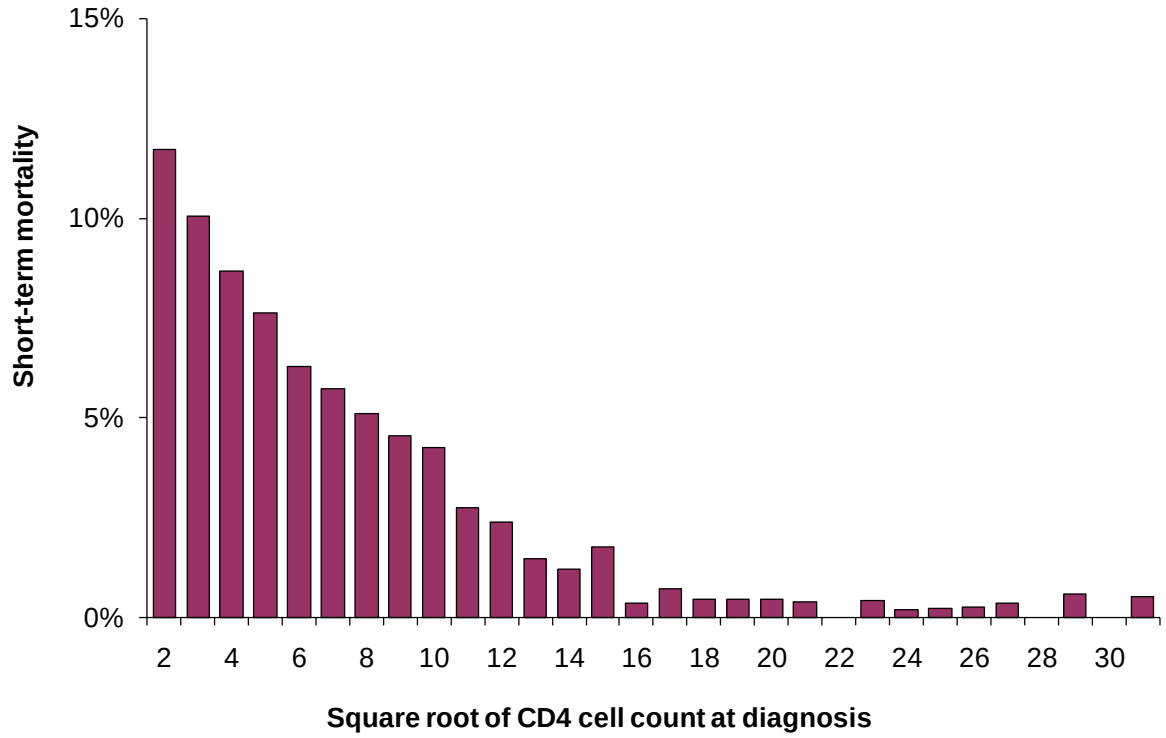


Figure 6.8. Short-term mortality and numbers diagnosed by square root of the CD4 cell count at HIV diagnosis (square root of CD4 less or equal to 31)



6.4.6 Late diagnosis and consequent mortality among heterosexuals with CD4 cell counts measured at diagnosis

In the integrated dataset, 19,882 heterosexuals were newly diagnosed with HIV infection between January 2000 and December 2007 and 17,843 (89.7%) of these individuals had a CD4 cell count measured within 91 days of their HIV diagnosis. Records were excluded from the analysis if they had evidence of previous diagnosis or evidence of recent seroconversion (9.7% [1,726] and 2.9% [524] respectively, of those with CD4 cell counts at diagnosis).

6.4.6.1 Patient characteristics

The majority (64.4%, 10,049 of 15,593) of newly diagnosed heterosexuals were female of which 19.5% (1,964/10,049) were diagnosed antenatally (Table 6.10). The median age at diagnosis of women diagnosed antenatally (28 years) was lower than that of other women (33 years) and men (36 years). Almost three quarters (74.6%) of heterosexuals were black African, 13.3% were white and 4.4% were black Caribbean. Half (49.2%) of white individuals were women compared to 43% of Indian/Pakistani/Bangladeshi individuals and 56.9% of black Caribbean individuals and two-thirds of all other ethnic groups. However, similar proportions of women were diagnosed through antenatal testing across ethnic groups. The median age at diagnosis was similar across ethnic groups when split by how infection was acquired except that white men were older than other men.

Where information on how infection was acquired was known, almost all (93.7% [10,422/11,126]) black Africans were probably infected in Africa (40% [4,170] of these in Zimbabwe) and 99.1% (8,632) of these were born in Africa (of the 8,714 individuals with a country of birth reported). Of these black Africans born and infected in Africa, 30.9% (2.666) had been in the UK less than two years prior to diagnosis, 20.0% (1,728) two or three years, 20.7% (1,785) four or more years and 28.4% (2,453) had no year of arrival reported. In contrast, two thirds (66.1% [466/705]) of white individuals infected outside the UK and with a reported country of birth were born in the UK (82.0% [382] of whom were men and 34.3% [160] of whom were white men infected in Thailand). Where information on how infection was acquired was known, almost half (45.1% [309/626]) of black Caribbeans were infected in the Caribbean (73.8% [228] of whom in Jamaica) and 24.2% were infected in the UK by a partner infected outside Europe. Half of black Caribbeans infected in the Caribbean were male (46.6% [144/309]) whereas 68.1% [113/166] of those infected in the UK by a partner infected outside Europe were male. Of black Caribbeans with a reported country of birth, 69.5% [232/334] of those infected outside the UK were born in the Caribbean whereas 46.8% [73/156] of those infected in the UK were born in the Caribbean.

Table 6.10. Characteristics of heterosexuals at HIV diagnosis by ethnicity and sex/reason for test

		Ethnicity							Total
		Black African	**White**	**Black Caribbean**	**Black Other**	**Indian\ Pakistani\ Bangladeshi**	**Other Asian**	**Other/mixed**	
Number diagnosed (percentage)	Male	3,754 (67.7)	1,052 (19.0)	295 (5.3)	121 (2.2)	118 (2.1)	69 (1.2)	135 (2.4)	5,544 (100)
	Female (not diagnosed antenatally)	6,312 (78.1)	837 (10.4)	315 (3.9)	213 (2.6)	70 (0.9)	133 (1.6)	205 (2.5)	8,085 (100)
	Female (diagnosed antenatally)	1,567 (79.8)	181 (9.2)	75 (3.8)	42 (2.1)	19 (1.0)	40 (2.0)	40 (2.0)	1,964 (100)
Median age at diagnosis (IQR)	Male	36 (31, 41)	41 (34, 51)	36 (30, 44)	35 (30, 38)	38 (32, 47)	34 (30, 40)	35 (30, 41)	36 (31, 43)
	Female (not diagnosed antenatally)	33 (28, 38)	33 (25, 42)	33 (28, 42)	31 (26, 39)	35 (29, 45)	29 (26, 34)	31 (26, 38)	33 (27, 39)
	Female (diagnosed antenatally)	28 (25, 32)	26 (22, 31)	28 (24, 31)	29 (23, 32)	25 (22, 28)	28 (25, 31)	29 (25, 33)	28 (24, 32)
Percentage infected abroad where known (%)	Male	92	58	64	70	79	77	81	83
	Female (not diagnosed antenatally)	90	28	57	58	49	80	78	80
	Female (diagnosed antenatally)	91	23	68	69	26	88	70	82

6.4.6.2 Late diagnosis

Two-fifths (41.8% 6,515 of 15,593) of heterosexuals were diagnosed late between 2000 and 2007 (Table 6.11). There were no significant differences in the proportions diagnosed late among individuals infected abroad. However, heterosexuals infected from a high-risk partner (30.0% diagnosed late; AOR 0.70; 95% CI 0.52, 0.95) and heterosexuals infected in the UK by partners infected outside Europe (23.3% diagnosed late; AOR 0.44; 95% CI 0.38, 0.51) were less likely to be diagnosed late than heterosexuals infected in Africa (44.4% diagnosed late). Women diagnosed antenatally (22.1% diagnosed late; AOR 0.49; 95% CI 0.44, 0.55) were less likely to be diagnosed late than other women (42.1% diagnosed late), who were less likely to be diagnosed late than men (48.3% diagnosed late; AOR 1.10; 95% CI 1.02, 1.19). Fewer individuals were diagnosed late in 2003/4 (40.5% [2,054/5,071]; AOR 0.83; 95% CI 0.72, 0.95) in comparison to 2000 (46.3% [493/1,064]) but there was no further decline to 2005/6 or 2007 (41.6% [1,854/4,459] and 40.8% [610/1,495] respectively). The percentage diagnosed late increased with age from 20.1% among those aged 15-25 years to 54.2% among those aged older than 54 years (AOR 0.33, 95% CI 0.29, 0.38 and AOR 1.38, 95% CI 1.15, 1.67 respectively compared to individuals aged 35-39 years). Individuals diagnosed outside London were less likely to be diagnosed late than those diagnosed in London (AOR 0.84; 95% CI 0.78, 0.90). A lower percentage of white (34.5% [714/2,070]; AOR 0.72; 95% CI 0.63, 0.83) and black Caribbean (35.3% [242/685]; AOR 0.72; 95% CI 0.63, 0.83) heterosexuals were diagnosed late than black Africans (44% [5,015/11,633]).

Among black African heterosexuals born and infected in Africa (those with clinician reports only) the length of stay in the UK prior to diagnosis was associated with late diagnosis: 42.5% (2,091/4,926) of those who had been in the UK less than five years prior to diagnosis were diagnosed late in contrast to 53.2% (667/1,253) of those who had been in the UK for five or more years prior to diagnosis ($p<0.01$); 44.0% (1,080/2,453) of those with no year of arrival reported were diagnosed late.

Table 6.11. Late diagnosis and factors associated with late diagnosis among heterosexuals with CD4 cell counts within 91 days of their HIV diagnosis

	Percentage of patients diagnosed late (CD4<200 cells/mm^3) % (n)	Multivariable adjusted odds ratio for late diagnosis (95% confidence interval)
How infection was acquired		
Heterosexual sex with a high-risk partner	30.0 (73)	**0.70 (0.52, 0.95)**
Heterosexual sex abroad:		
In Africa	44.4 (5,017)	-
In Latin America/Caribbean	40.0 (179)	0.96 (0.75, 1.23)
An Asia	45.9 (272)	0.90 (0.72, 1.12)
In other region abroad	37.5 (125)	0.90 (0.70, 1.16)
Heterosexual sex in the UK:		
With a partner infected outside Europe	23.3 (310)	**0.44 (0.38, 0.51)**
With a partner infected within Europe	42.5 (113)	1.08 (0.82, 1.42)
With a partner infected in an unknown location	34.3 (60)	0.74 (0.53, 1.04)
Heterosexual sex: no further information	40.9 (366)	**0.79 (0.67, 0.92)**
Sex/reason for test		
Male	48.3 (2,675)	**1.10 (1.02, 1.19)**
Female (not diagnosed antenatally)	42.1 (3,405)	-
Female (diagnosed antenatally)	22.1 (435)	**0.49 (0.44, 0.55)**
Period of HIV diagnosis		
2000	46.3 (493)	-
2001-2002	42.9 (1,504)	0.89 (0.77, 1.03)
2003-2004	40.5 (2,054)	**0.83 (0.72, 0.95)**
2005-2006	41.6 (1,854)	**0.87 (0.75, 1.00)**
2007	40.8 (610)	**0.84 (0.70, 0.99)**
Age group at HIV diagnosis		
15-25	20.1 (373)	**0.33 (0.29, 0.38)**
25-29	32.0 (1,040)	**0.56 (0.50, 0.62)**
30-34	42.5 (1,575)	**0.83 (0.75, 0.91)**
35-39	48.3 (1,421)	-
40-44	54.6 (947)	**1.29 (1.14, 1.45)**
45-49	55.2 (550)	**1.36 (1.17, 1.57)**
50-54	55.3 (301)	**1.42 (1.17, 1.71)**
>54	54.2 (308)	**1.38 (1.15, 1.67)**
Region of diagnosis		
Outside London	39.6 (3,229)	**0.84 (0.78, 0.90)**
London	44.2 (3,286)	-
Ethnicity		
Black African	43.1 (5,015)	-
White	34.5 (714)	**0.72 (0.63, 0.83)**
Black Caribbean	35.3 (242)	**0.76 (0.61, 0.94)**
Black Other	42.8 (161)	1.09 (0.87, 1.36)
Indian\Pakistani\Bangladeshi	44.4 (92)	1.02 (0.75, 1.40)
Other Asian	49.2 (119)	**1.70 (1.24, 2.31)**
Other/mixed	45.3 (172)	1.17 (0.94, 1.46)
Total	41.8 (6,515)	-

6.4.6.3 Short-term mortality

The overall short-term mortality of heterosexuals with CD4 cell counts within 91 days of diagnosis was 2.1% (329/15,593) (Table 6.12). The majority (87.2% [287/329]) of those who died within a year of diagnosis were diagnosed late. Hence, the short-term mortality was 4.4% (287/6,515) among those diagnosed late and 0.5% (42/9,078) among others (AOR 7.66, 95% CI 5.51, 10.67).

Independently of late diagnosis, short-term mortality was higher for heterosexuals infected in the UK by partners infected within Europe (AOR 2.25; 95% CI 1.14, 4.44) than for heterosexuals infected in Africa. Women diagnosed antenatally (0.2% [4/1,964] died within a year of diagnosis; AOR 0.21; 95% CI 0.08, 0.56) were less likely to experience short-term mortality than other women (1.9% [151/8,085] died within a year of diagnosis), who were less likely to experience short-term mortality than men (3.1% [174/5,544] died within a year of diagnosis; AOR 1.35; 95% CI 1.07, 1.70). Older age was associated with short-term mortality (AOR 2.67; 95% CI 1.72, 4.15 for those aged older than 54 years at diagnosis compared to those aged 35-39 years). Individuals of 'black-other' ethnicity were the only ethnic group to be less likely to die within a year of diagnosis than black Africans after controlling for other factors (AOR 0.33; 95% CI 0.13, 0.84). There was no significant regional effect (AOR 1.01; 95% CI 0.80, 1.28).

Table 6.12. Short-term mortality and factors associated with short-term mortality among heterosexuals with CD4 cell counts within 91 days of their HIV diagnosis

	Short-term mortality among heterosexuals with CD4 cell counts at HIV diagnosis		Multivariable adjusted odds ratio for short-term mortality (95% confidence interval)
	CD4 < 200 % (n)	CD4 ≥ 200 % (n)	
Late HIV diagnosis status			
CD4 < 200	4.4 (287)	-	**7.66 (5.51, 10.67)**
CD4 ≥ 200	-	0.5 (42)	-
How infection was acquired			
Heterosexual sex with a high-risk partner	4.1 (3)	0 (0)	0.69 (0.21, 2.34)
Heterosexual sex abroad:			
In Africa	4.3 (214)	0.6 (37)	-
In Latin America/Caribbean	5.6 (10)	0.4 (1)	1.95 (0.90, 4.21)
An Asia	4.4 (12)	0.3 (1)	0.71 (0.34, 1.45)
In other region abroad	7.2 (9)	0.5 (1)	1.27 (0.61, 2.66)
Heterosexual sex in the UK:			
With a partner infected outside Europe	3.5 (11)	0.1 (1)	0.67 (0.36, 1.25)
With a partner infected within Europe	11.5 (13)	0.7 (1)	**2.25 (1.14, 4.44)**
With a partner infected in an unknown location	8.3 (5)	0 (0)	1.27 (0.47, 3.42)
Heterosexual sex: no further information	2.7 (10)	0 (0)	**0.51 (0.26, 0.99)**
Sex (reason for test)			
Male	5.6 (150)	0.8 (24)	**1.35 (1.07, 1.70)**
Female (not diagnosed antenatally)	4.0 (135)	0.3 (16)	-
Female (diagnosed antenatally)	0.5 (2)	0.1 (2)	**0.21 (0.08, 0.56)**
Period of HIV diagnosis			
2000	4.5 (22)	1.9 (11)	-
2001-2002	4.7 (71)	0.4 (8)	0.80 (0.52, 1.22)
2003-2004	4.2 (87)	0.4 (13)	0.76 (0.50, 1.15)
2005-2006	4.5 (84)	0.3 (9)	0.77 (0.51, 1.18)
2007	3.8 (23)	0.1 (1)	0.61 (0.35, 1.07)
Age group at HIV diagnosis			
15-25	1.9 (7)	0.1 (2)	0.49 (0.24, 1.00)
25-29	3.2 (33)	0.3 (6)	0.86 (0.57, 1.30)
30-34	3.0 (48)	0.5 (11)	0.85 (0.59, 1.22)
35-39	3.7 (53)	0.8 (12)	-
40-44	4.9 (46)	0.5 (4)	1.16 (0.79, 1.69)
45-49	7.5 (41)	1.1 (5)	**1.94 (1.31, 2.88)**
50-54	7.6 (23)	0.8 (2)	**1.85 (1.14, 3.01)**
>54	11.7 (36)	0 (0)	**2.67 (1.72, 4.15)**

Region of diagnosis				
	Outside London	4.5 (146)	0.5 (23)	1.01 (0.80, 1.28)
	London	4.3 (141)	0.5 (19)	-
Ethnicity				
	Black African	4.1 (208)	0.5 (33)	-
	White	7.3 (52)	0.4 (6)	1.05 (0.67, 1.64)
	Black Caribbean	2.9 (7)	0 (0)	**0.33 (0.13, 0.84)**
	Black Other	3.7 (6)	0.9 (2)	1.05 (0.50, 2.21)
	Indian\Pakistani\Bangladeshi	4.3 (4)	0 (0)	0.77 (0.26, 2.26)
	Other Asian	4.2 (5)	0.8 (1)	1.54 (0.58, 4.05)
	Other/mixed	2.9 (5)	0 (0)	0.54 (0.21, 1.37)

6.4.7 Crude estimation of preventable short-term mortality

There were 10,787 MSM diagnosed between 2000 and 2007 with no evidence of previous diagnosis. Among the 9,742 with CD4 cell counts within 91 days of diagnosis, 22.3% were diagnosed late. Short-term mortality among those diagnosed late was 4.8% compared to 0.4% among those diagnosed with higher CD4 cell counts. Assuming that those with CD4 cell counts at diagnosis were representative of all MSM diagnosed between 2000 and 2007, an estimated 149 MSM would have died within a year of HIV diagnosis with 115 (77.5%) of them being diagnosed late. Assuming that earlier diagnosis and treatment could have reduced the mortality rate among MSM diagnosed late to that among MSM diagnosed promptly, only 43 deaths would have been observed within a year of HIV diagnosis with 10 (22.3%) of them among those that would have been diagnosed late. The estimated 105 preventable deaths accounted for 71.0% of all estimated deaths within a year of diagnosis and 11.1% (115/952) of all deaths that occurred among MSM between 2000 and 2007.

There were 13,316 black African heterosexuals, 2,475 white heterosexuals and 827 black Caribbean heterosexuals diagnosed between 2000 and 2007 with no evidence of previous diagnosis or seroconversion (Table 6.13). Assuming that late diagnosis and consequent short-term mortality was not preventable among recent arrivals in the UK, there would have been an estimated 96 deaths within a year of HIV diagnosis among black Africans recently arrived in the UK and diagnosed late. No deaths would have been expected among white individuals recently arrived and diagnosed late and only four would have been expected among black Caribbeans recently arrived and diagnosed late. Assuming that short-term mortality among other individuals could be reduced to that of individuals who were not diagnosed late by preventing late diagnoses, 55, ten and zero deaths within a year of diagnosis would have been expected among these three ethnic groups. The combined number of deaths expected within a year of diagnosis was 60% (151/252) of that observed for black African heterosexuals, 50% (4/8) for black Caribbean heterosexuals and only 17% (10/60) for white heterosexuals. The potentially preventable deaths accounted for 18% (101/560), 13% (4/30) and 23% (50/217) of all deaths observed between 2000 and 2007 for black African, black Caribbean and white heterosexuals respectively.

Numbers were large enough to analyse preventable short-term mortality among black African heterosexuals by sex/reason for test. An estimated 78 deaths within a year of diagnosis would have been expected among men, of which 48 (61%) would have been among recent arrivals diagnosed late and 30 (39%) would have been among other individuals. Prevention of late diagnosis among

non-recent arrivals could have reduced the observed number of deaths within a year of diagnosis among men by 35% (from 120) and the number of all deaths by 16% (from 255 to 213) during the period. A similar effect would have been expected among women: an estimated 77 deaths within a year of diagnosis would have been expected – a 41% decrease in short-term mortality and an 18% decrease in all mortality during the period. Mortality was so low for women diagnosed antenatally that the estimated number of deaths within a year of diagnosis was actually greater than the number observed indicating that there was no demonstrable preventable short-term mortality among this group.

Estimated preventable short-term mortality increased slightly but did not change markedly if heterosexuals without a reported country of birth or year of arrival were assumed to have been in the UK at least two years prior to HIV diagnosis (although numbers were small for black Caribbean heterosexuals).

Table 6.13. Estimation of preventable short-term mortality among heterosexuals.

	Number of individuals newly diagnosed	Proportion of recent arrivals	Proportion of recent arrivals diagnosed late	Estimated short-term mortality among recent arrivals diagnosed late	Estimated short-term mortality among other individuals if not diagnosed late	Potential short-term mortality as fraction of observed short-term mortality	Potentially preventable short-term mortality	Potentially preventable short-term mortality as a percentage of all deaths
Assuming heterosexuals with a reported country of birth and year of arrival were representative of all heterosexuals								
All black African heterosexuals	13,316	41% (5,460)	42% (2,285)	4.2% (96)	0.5% (55)	151 / 252 (60%)	40% (117)	18% (117 / 560)
All white heterosexuals	2,475	4% (97)	33% (32)	0% (0)	0.4% (10)	10 / 60 (17%)	83% (50)	23% (50 / 217)
All black Caribbean heterosexuals	827	26% (215)	37% (80)	4.6% (4)	0% (0)	4 / 8 (50%)	50% (4)	13% (4 / 30)
Black African heterosexual men	4,289	40% (1,733)	52% (901)	5.3% (48)	0.9% (30)	78 / 120 (65%)	35% (42)	16% (42 / 255)
Black African heterosexual women (not diagnosed antenatally)	7,310	44% (3,221)	42% (1,353)	3.9% (53)	0.4% (23)	77 / 130 (59%)	41% (53)	18% (53 / 287)
Black African heterosexual women (diagnosed antenatally)	1,717	38% (655)	21% (134)	1.2% (2)	0.1% (1)	3 / 2 (150%)	-	-
Assuming heterosexuals without a reported country of birth or year of arrival were in the UK at least two years prior to HIV diagnosis								
All black African heterosexuals	13,316	29% (3,879)	42% (1,624)	4.2% (68)	0.5% (58)	127 / 252 (40%)	50% (125)	22% (125 / 560)
All white heterosexuals	2,475	5% (117)	33% (38)	0% (0)	0.4% (10)	10 / 60 (17%)	83% (50)	23% (50 / 217)
All black Caribbean heterosexuals	827	11% (91)	37% (34)	4.6% (2)	0% (0)	2 / 8 (25%)	75% (6)	20% (6 / 30)

6.5 Discussion

6.5.1 Overview

This analysis estimated the proportion of individuals diagnosed late and subsequent short-term mortality. It also identified groups that were more likely to be diagnosed late with HIV infection and more likely to die before treatment could improve or stabilise their condition. Between 1995 and 2007, almost two-fifths of all individuals were diagnosed late and one in fifty of all individuals died within a year of their HIV diagnosis. Overall, late diagnosis decreased over time, increased with age and was significantly lower for MSM, heterosexual women diagnosed antenatally, individuals of white ethnicity and individuals diagnosed in London than for other groups. Among heterosexuals diagnosed since 2000, there was evidence that individuals with a high-risk partner and individuals infected in the UK by a partner from outside Europe were less likely to be diagnosed late. Most importantly, the results show that although short-term mortality decreased over time independently of CD4 counts at diagnosis, and presumably due to HAART, those diagnosed late were still approximately ten times more likely to die within a year of diagnosis than comparable individuals who were not diagnosed late (5.4% versus 0.5%). The results indicate that substantial mortality could be avoided by reducing late diagnoses.

The decline in the proportion of HIV-infected individuals diagnosed late between 1995 and 2007 suggests that increases in HIV testing[366] have resulted in earlier diagnosis of HIV-infected individuals (there was no evidence from these analyses that the availability of HAART in 1996 resulted in a marked increase in earlier patient-initiated HIV testing). This means that each year a higher

proportion of those diagnosed with HIV could be offered therapy early and therefore expect a longer and healthier life[124;330;367;368]. However, the increase in numbers of diagnoses of black African heterosexuals (excluding women diagnosed antenatally) who were more likely to be diagnosed with low CD4 cell counts resulted in about a third of individuals still being diagnosed late since 2003 and substantial effort is still required to counter this situation. In addition to lives lost, there are financial implications of late diagnosis because direct care costs in the year following diagnosis are more than 200% higher for patients diagnosed late[331]. However, a much higher proportion of non-white heterosexuals were infected abroad[369], and therefore, late diagnoses among heterosexuals also reflects a lack of opportunities to diagnose earlier.

Older age, in particular, was associated with both late diagnosis and short-term mortality and therefore diagnosis earlier in the course of infection, and at a younger age, would synergistically reduce short-term mortality. This was evident in that the adjusted odds ratio for short-term mortality among women diagnosed antenatally was approximately one fifth of that among other women even after controlling for age and late diagnosis. Interestingly, the lower proportion of MSM diagnosed late at a similar age to heterosexual men and women (not diagnosed antenatally) suggests that they may have been infected later in life, which implies that prevention of transmission among heterosexuals may need to target heterosexuals at a younger age than MSM. This warrants further investigation before changing any public health interventions as this observation could also be a result of faster progression or lower pre-infection CD4 cell counts among heterosexuals.

Women and white individuals may have been infected longer than men and other ethnicities diagnosed with similar CD4 counts due to higher CD4 counts prior to diagnosis – if these differences persist over time (Section 1.3.2). However, proportions diagnosed late across sex and ethnic groups are comparable for risks of progression because time since infection, sex and ethnic group have little clinical significance on progression of infection after adjusting for CD4 cell count and other biomarkers/clinical information[49;79;81;82;370] (hence treatment guidelines do not recommend initiation of antiretroviral treatment at different CD4 counts according to sex and ethnicity).

Although late diagnosis was less common among MSM than other HIV-infected individuals[371;372], around a fifth were diagnosed late in 2007, which supports findings of a continued high proportion of late diagnoses among MSM[300;320]. This is likely to be due to men not presenting for or declining HIV testing because it is likely that most actively homosexual men in the UK are aware of the risk of HIV infection[373]. As most MSM newly diagnosed in the UK were infected in the UK and have been exposed to years of health promotion, this should strengthen calls for targeted messages to highlight the benefits of early testing and for HIV testing policy to target this issue.

Earlier diagnosis of MSM since 1995 may be due in part to increased awareness of the benefits of early testing due to HIV testing promotion efforts[374;375]. However, increasing incidence in MSM[376], increasing age at infection, and a decreasing contribution to new diagnoses of long-term undiagnosed MSM who were infected during the peaks in incidence of HIV in 1983 and in 1989/1990[377] could also account for part of this trend. Yet,

encouragingly, MSM coming forward earlier for testing were likely to have contributed to increases in the numbers of new HIV diagnoses since 2000[369].

The results suggest that groups at high risk of late diagnosis (including individuals born in the UK who were exposed to HIV in high prevalence countries) should be targeted for appropriate public health action. Such action might include health promotion, opportunistic screening, and removal of any barriers to testing. Although the proportion of MSM and sub-Saharan-born genitourinary medicine (GUM) clinic attendees who accepted HIV testing increased to 86% and 85% respectively in 2007, 35% of MSM and 30% of sub-Saharan-born individuals whose HIV infection could have been diagnosed during their GUM clinic attendance left undiagnosed[378;379]. Perceptions of risk, accessibility of healthcare, fear of deportation, fear of criminalisation and community stigma[171;297;353;380-385] may all reduce the success of testing policies. In particular, there has been concern that the exclusion of 'irregular' migrants, including failed asylum seekers and visa-overstayers, from eligibility to free HIV treatment from the NHS created a barrier to HIV testing, and may have hindered efforts to reduce transmission[386;387]. This may have been compounded by social and economic insecurity among 'irregular' migrants whose life circumstances may consequently place them at greater risk of HIV infection.

Even if some undiagnosed HIV-infected individuals are encouraged to seek a diagnosis through targeted HIV testing promotion campaigns, opportunistic screening in a variety of healthcare settings should be considered and among groups other than MSM, pregnant women and GUM clinic attendees[106-109;342;388-392]. This could be cost-effective even without considering the benefits

of prevented transmission[393]. The intensification of antenatal HIV screening, move towards opt-out HIV testing in GUM clinics and change from pre-test counseling to pre-test discussion have been very successful[115;394;395]. Data are now available to support the geographical targeting of increased HIV testing[396] but robust evaluation of HIV testing in a variety of settings in the locality of communities at high risk of HIV infection is now required. In particular, new patient checks in primary care may provide the earliest opportunity to diagnose infection among new entrants to the UK[107;297;385]. However, seeking an HIV test may not be the highest priority for these individuals who may also face financial, housing, immigration, relationship and discrimination problems in the UK[236]. Short-term mortality may also be reduced by increasing awareness of the risks associated with late diagnosis among professionals carrying out health assessments for asylum seekers from Africa and the Caribbean (these professionals are already recommended to consider HIV[397;398]). This is suggested given the markedly increased numbers of applications for asylum from Zimbabweans and Jamaicans between 1999 and 2002[399;400] and the large number of HIV-infected heterosexuals with insecure immigration or asylum seeker status[401;402].

Late diagnosis also varied by region of diagnosis, which may be due to factors that could not be considered, such as differences between populations (for example awareness of the benefits of HIV testing) or differences in testing policy and practice (for example contact tracing and partner notification). It was likely to be due to a combination of community and provider characteristics as MSM were diagnosed earlier than heterosexuals and because HIV testing uptake is known to vary between GUM and antenatal clinics[115].

6.5.2 Preventable mortality

Estimated short-term mortality accounted for over a quarter of all deaths between 2001 and 2007 and was ten times higher among newly diagnosed individuals who were diagnosed late than among those diagnosed promptly. The scope for reducing late diagnoses and the consequent short-term mortality of recent arrivals to the UK may be limited, as they may already have low CD4 counts before arrival in the UK. There were marked differences in the proportion of recent arrivals between ethnic groups with nine-tenths of black Africans being infected abroad in particular. Therefore, less short-term mortality could have been prevented by earlier diagnosis in the UK among black African and black Caribbean heterosexuals (short-term mortality could have been reduced by 40% and 50% respectively) than was possible among MSM and white heterosexuals (short-term mortality could have been reduced by about four-fifths). Yet, the absolute number of preventable deaths among black African heterosexuals over the period and the potential reduction in overall mortality may have been larger than that among MSM because estimated short-term mortality accounted for a higher proportion of all deaths (45% (252/560) of all deaths among black African heterosexuals occurring between 2000 and 2007 versus 16% (149/952) of all deaths among MSM and 28% (60/217) of all deaths among white heterosexuals). This analysis suggests that reductions in short-term mortality resulting from prompt diagnosis of all individuals could have given a 11% (105/952), 18% (101/560) and 23% (50/217) reduction in the number of all deaths among MSM, black African heterosexuals and white heterosexuals, respectively, between 2000 and 2007 (deaths among black Caribbeans were too few to calculate robust estimates). Numbers of deaths among black Africans

were large enough to suggest that less short-term mortality may have been preventable among men (35%) than women who were not diagnosed antenatally (41%). Given that many individuals diagnosed late had been in the UK for two or more years prior to diagnosis, the potential for reducing short-term mortality by earlier diagnosis is substantial. It is even possible that significant reductions in mortality could be achieved among recent arrivals by earlier diagnosis even if these CD4 cell counts were not greater than 200 cells/mm^3. Pre-treatment deaths and deaths that occur shortly after starting HAART could be prevented if diagnosis was made earlier in the course of infection.

6.5.3 Comparisons

6.5.3.1 Late diagnosis

The percentage of HIV-infected individuals diagnosed late in other developed countries varied substantially[168;339;372;391;403-411] and trends over time also varied[405-407;409;410;412]. However, the factors associated with late diagnosis indicate that comparisons between countries should consider the stage of the HIV epidemic and trends in diagnoses among different risk groups[410].

Other studies have reported associations between population factors and the late diagnosis of HIV infection similar to those found in this analysis: age[271;301;405;407-410;413;414]; sex[301;408-410;413;414]; risk group[301;372;405;407;408]; ethnicity/race[271;410;413;415]; time[301;406;408;415]; migration[301;391;409]. Some studies have identified factors associated with late diagnosis that were not available in the integrated dataset. These include being unemployed, in a couple or having children[408], a lack of earlier/routine testing[405;410] and diagnosis at health care

settings[405;410] (the latter is an effect of, rather than a possible determinant of, late diagnosis and so would not help identify groups at increased risk).

Finally, comparisons between reports of late diagnosis should consider the setting in which the studies were conducted because results and changes over time may be locally specific[416]. This may explain why data from a subset of larger UK clinics showed that 26.8% of individuals were diagnosed late[338].

6.5.3.2 Consequent mortality

The ART Cohort Collaboration showed that the probability of death in the first year of treatment was lower for individuals who started HAART with a CD4 count greater than 199 cells/mm^3 than for those who started HAART late[124]. The higher mortality in the era of HAART availability among individuals who were diagnosed late in my analyses (4.1% of non-IDU individuals aged younger than 50 years died within a year and 10.6% of non-IDU individuals aged 50 years or older) may be due to a high number and proportion of deaths occurring in those diagnosed late between diagnosis and treatment initiation. I aimed to investigate the consequences of late diagnosis rather than the consequences of late initiation of ART and suggest that there are many deaths that could be prevented by earlier diagnosis. Similar results have subsequently been found elsewhere[412;417]. Reviews of deaths[418] and hospitalisations[419] among HIV-infected individuals in the UK found that 24% and 17% were due to late diagnosis and that this was the major preventable cause of death[418]. Late diagnosis and consequent mortality can be prevented with repeated HIV testing and effective linkage to, and retention in, care [168;271;413;414;420-422].

6.5.3.3 Late diagnosis of heterosexuals

Other studies in London have found that black-Africans tend to present later than white individuals[80;321;324;326]. However, these have not controlled for differences between risk groups or how/where infection was acquired, which explained differences in the proportion diagnosed late between ethnicities in this analysis. Increasing age and male sex were consistently associated with late diagnosis in studies from other countries but these did not consider heterosexuals separately and did not show a difference between heterosexuals and homosexual men[168;331;423]. However, two of these studies also considered HIV testing history and aspects of sexual behaviour, which were likely to be associated with sexual orientation.

Evidence of earlier diagnosis among individuals with a high-risk partner and individuals infected in the UK by a partner from outside Europe suggests that awareness of risk among these groups resulted in them seeking healthcare and requesting HIV tests. Earlier diagnosis among women diagnosed antenatally supports evidence that provider-initiated offer of HIV testing is effective.

6.5.4 Limitations

There are limitations to these analyses. The lack of information about viral loads at HIV diagnosis may limit the determination of odds ratios of short-term mortality because viral load has been shown to be predictive of progression of infection independently of CD4 counts and age for treatment naïve individuals [299;424]. However, in patients starting HAART, CD4 cell counts are the dominant prognostic factor and viral load levels have relatively little effect on subsequent

disease progression[124;361]. The effect of adherence could not be addressed as surveillance does not collect this information.

The dataset contains some missing data, in particular, of where/how infection was acquired and year of arrival. How/where infections were acquired was a significant factor associated with late diagnosis but was not available in the majority of reports of diagnoses among MSM requiring this information to be analysed separately. Lack of information on year of arrival for the majority of MSM precluded the consideration of this factor in estimation of preventable mortality in this group. Furthermore, women diagnosed antenatally may not all have been identified through reports, which may have contributed to the lower proportion of other women diagnosed late compared to men.

Despite the large overall number of records analysed, small numbers for all groups except black-African heterosexuals born and infected in Africa limited comparisons when stratified by other factors and the conclusions that could be drawn from these data.

Further reports were expected of deaths within a year of diagnosis and individuals diagnosed towards the end of 2007 with CD4 cell counts within 91 days of diagnosis may not have been included because those measurements occurred in 2008. These biases due to censoring of the data were particularly likely to affect the results of individuals who were diagnosed in 2007 but were unlikely to change the overall results.

For the generalisation of these results to all new diagnoses in E,W&NI the immunological status of individuals with known CD4 counts should be

representative of comparable individuals without known CD4 counts. There was therefore significant potential for bias from not including those without CD4 cell counts reported to either surveillance system and those with CD4 cell counts more than 91 days after diagnosis.

There were significant differences between groups in the proportions included in the integrated dataset (Sections 5.4.1 and 5.5.1). Heterosexuals, IDU, non-white individuals, older people and individuals diagnosed outside London were under-represented in the integrated dataset and, as all of these groups were associated with late diagnosis, it was likely that the overall percentage diagnosed late, and therefore consequent mortality, were under-estimated. Consequent mortality was also likely to be substantially under-estimated because individuals who died were less likely to be included in the integrated dataset and because individuals who died within a month of HIV diagnosis were less likely to be included than other individuals who died (Sections 5.4.1 and 5.5.1). As those who died and were not included in the integrated dataset were likely to have low CD4 cell counts, the adjusted odds ratios and statistical significance for late diagnosis were also likely to be underestimated. Therefore, heterosexuals (not diagnosed antenatally), IDU, non-white individuals, older people and individuals diagnosed outside London were even more likely to have been diagnosed late than the results indicate. Groups least likely to be included in the integrated dataset (such as individuals aged 55 years or older, IDU, black Africans and black – other individuals, and individuals diagnosed outside London) would be more likely to have truly higher odds of late diagnosis.

The analysis does not reflect late entry to clinical care because only individuals with CD4 cell counts within 91 days of diagnosis were included. Furthermore, about a tenth of records included in the integrated dataset were excluded due to evidence of previous diagnosis. Some of these individuals will only have had a previous positive HIV test result and not previously entered clinical care.

Records with evidence of recent infection were also excluded. However, some of the 8.6% individuals observed with CD4 cell counts less than 200 cells/mm^3 would not have experienced spontaneous immune reconstitution after seroconversion. These individuals could be considered as diagnosed late with regards their prognosis but probably could not have been diagnosed significantly earlier to change their prognosis or reduce transmission[343].

Varying the definition of late diagnosis had little effect on the associations with the determining factors. Increasing the CD4 cell count cut-off to 350 cells/mm^3 (to reflect treatment guidelines published in 2008[123]) and including AIDS in the definition of late diagnosis increased the proportions diagnosed late but similarly in most groups. The use of AIDS-defining illnesses to define late diagnoses would be likely to introduce bias in this study. This is because only individuals reported by clinicians can be identified as having AIDS and clinicians may be likely to report if an individual has AIDS at HIV diagnosis. Additionally, the prevalence of AIDS defining illnesses is likely to vary among populations and the CD4 cell counts at which these occur varies[355;356;425], which means that proportions diagnosed late defined by AIDS diagnoses may not be comparable between studies, countries or over time. Requiring two CD4 cell counts to define late diagnosis increased or decreased the proportion diagnosed late

depending upon the methodology but again, this had relatively little impact on the results.

Late diagnosis and its impact on short-term mortality may have been underestimated if HAART had significantly increased CD4 counts between diagnosis and the first CD4 count. However, almost a quarter of individuals involved in this analysis had CD4 counts within a day of the date of HIV diagnosis, and a large majority within a month. HAART is unlikely to be started before blood is taken for determination of the CD4 cell count and the delay between diagnosis and first CD4 cell count is likely to be accounted for by confirmation of infection (there are no national guidelines for how this process should be managed). There were significant differences between groups with regard the proportions with CD4 cell counts within 91 days of HIV diagnosis. Sensitivity analyses, which considered different cut-offs of the time between diagnosis and first CD4 cell count in the definition of late diagnosis, showed that some of these differences were artificial. The cut-off of 91 days was judged to be the best balance between maximising inclusion of records and minimising inclusion of CD4 cell counts that may have changed since diagnosis (for optimal representativeness). However, there was an association between having AIDS at diagnosis and having CD4 cell counts at diagnosis, which may, in particular, indicate confounding and suggest that late diagnosis was over-estimated overall and that odds ratios were over-estimated for groups that were more likely to be diagnosed with AIDS. The greater proportion of CD4 cell counts among individuals with AIDS at HIV diagnosis may reflect prioritisation of measuring CD4 cell counts among people diagnosed with AIDS. However, around three-fifths of all individuals diagnosed in E,W&NI between 1995 and 2007 were

represented in this analysis and bias should be minimised because deaths and CD4 cell counts are independently collected and matched to reports of new HIV diagnoses. Information missing due to inability to link records may be more likely to be missing at random than if solely collected through clinician reports. Additionally, the similar distributions of CD4 cell counts reported to CD4 Surveillance, and CD4 cell counts reported by clinicians to HARS that were not available from CD4 Surveillance, supported the hypothesis that missing CD4 cell counts were not markedly different from those analysed.

Varying the definition of short-term mortality by using different cut-off times from HIV diagnosis and death significantly changed some of the comparative odds ratios. In particular, increased cut-off time between diagnosis and death was reflected in a decreased magnitude of the effect of late diagnosis on short-term mortality and a corresponding increase in the effects of period of HIV diagnosis, age group and risk group. This indicates that the effect of late diagnosis on mortality diminishes over time as is clear from trends in the number of deaths by time from HIV diagnosis. Increasing the cut-off time to death in the analysis increasingly took into account factors that were associated with death post-HIV diagnosis, explaining the increase in the adjusted odds ratio for IDU. However, limiting the definition of short-term mortality to deaths within three months of diagnosis would not capture deaths due to late diagnosis that occurred later in the year and therefore would have underestimated the effect of late diagnosis.

Consideration of the association of CD4 cell count categories at the time of HIV diagnosis with short-term mortality indicated that there was not a clear difference between diagnosis with CD4 cell counts slightly less than 200

cells/mm^3 or slightly more. There were marked declines in short-term mortality as CD4 cell counts increased to 250-299 cells/mm^3 but then little further decline thereafter. Short-term mortality was lower among individuals diagnosed with CD4 cell counts between 150-199 cells/mm^3 than at 200-249 cells/mm^3 although this was not statistically significant. This may reflect selection by indication[426], in that individuals diagnosed with CD4 cell counts less than 200 cells/mm^3 were more promptly started on ART according to guidelines[135] with consequently lower short-term mortality than those with CD4 cell counts between 200-249 cells/mm^3. This analysis suggests that individuals would be particularly more likely to survive the first year after diagnosis if they were diagnosed before their CD4 cell counts fell below 250 cells/mm^3. Guidelines released in 2008 recommend that individuals start ART as soon as they are ready after CD4 cell counts fall below 350 cells/mm^3 [123], supported by evidence from the SMART clinical trial and previous cohort studies[427-429]. This is more conservative than my conclusion but none of these results can specifically support the cost-effectiveness of the guideline threshold as the risk of progression appears to fall with increasing CD4 cell counts. However, this has brought UK guidelines into conformity with European[362] and American guidelines[430] and has raised calls for harmonisation of a definition of late diagnosis as CD4 cell counts less than 350 cells/mm^3 [337-339;342].

6.5.5 Further work not presented in this thesis

Further work was carried out in the development of these analyses that is not presented in this thesis[239]. Inverse probability weighting was used to account for missing observations: the CD4 cell counts that were not available at the time of HIV diagnosis. This enabled the estimation of numbers diagnosed late and the

number of deaths within a year of diagnosis assuming that those with CD4 cell counts at the time of HIV diagnosis were representative, in these regards, of all individuals newly diagnosed. The estimates produced were in close agreement with observed short-term mortality for all individuals newly diagnosed. This methodology could be particularly useful for analysing surveillance data, which may be incomplete due to under-reporting or due to limitations of matching datasets, but where the characteristics of the total population, including which individuals are included in 'nested' analyses, are known[431].

6.6 Conclusion

Low CD4 cell counts at diagnosis reflect in general a long time from infection and are often indicative of a low awareness of individual risk of infection, a low personal drive to know one's own status and/or a lack of encouragement from health services to be tested. Low CD4 counts also indicate that individuals have been unaware for a long time that they might infect their partner/s and unaware that they could reduce this risk through behavioural change[32;334;363;432;433]. Health promotion efforts can aim to improve awareness of risks, knowledge of access to health services, understanding of the benefits of knowing one's HIV status and the acceptability of having an HIV test. HIV testing strategies can aim to improve the accessibility of HIV tests by offering and encouraging testing at a variety of health services in a culturally sensitive and acceptable fashion.

In addition to the benefits to patients of an early diagnosis, there are substantial public health benefits associated with reducing the prevalence of undiagnosed HIV infection and reducing transmission. Almost 30% of HIV-infected individuals were estimated to be living with undiagnosed infection in 2007 (over two-fifths of

HIV-infected non-African men[394]). The Department of Health's National Strategy for Sexual Health and HIV for England[434] in 2001 aimed to reduce the prevalence of undiagnosed infection. While estimates suggest that it may have fallen among heterosexuals from 42% in 2001 to 30% in 2007, more needs to be done, particularly as the undiagnosed fraction remained stable about 25% among MSM over the same period[394;435]. Large numbers remain undiagnosed for a long time, and particularly those who are diagnosed late with high viral loads, may be a major source of new infections[436]. As well as not benefiting from therapy they form a reservoir for potential unwitting transmission and cannot take measures to reduce the risk of transmission to their partners by using HAART or changing their behaviour[432;433;437;438]. Both the National Strategy for Sexual Health and HIV for England[434] and the Chief Medical Officer for England's 2004 annual report[333], prioritised the uptake of HIV testing in a variety of healthcare settings as a core HIV prevention intervention. This has also been a recommendation in other countries since 2003[439]. This strategy aims to reduce the number of missed diagnoses[423;440-444] and numbers diagnosed as hospital inpatients[236;297;419;444;445], thus reducing costs[331]. Development of HIV testing services in general medical practices, including for individuals from high prevalence countries[388;446;447], may reduce late diagnosis for HIV-infected heterosexuals who also may be less likely to access other health care services[171;323]. In addition to testing, further resources may also need to be committed to contact tracing and partner notification to identify infected individuals early and reduce onward transmission, particularly in low prevalence areas. Clinicians should be particularly aware of the likely need for prompt initiation of HAART among men, older individuals, individuals from ethnic minorities or infected abroad and individuals diagnosed outside London.

Chapter 7. National incidence rates of death and first AIDS diagnoses

7.1 Introduction

Surveillance data are used to monitor trends in the numbers of AIDS diagnoses and deaths but the public health utility of this information can be increased by considering the number of people and characteristics of people at risk of these events with the integrated dataset. HARS data show that numbers of AIDS cases diagnosed and deaths were fairly stable between 1997 and 2007 whereas SOPHID shows that the number of people living with diagnosed HIV infection increased during this same time period. Crude comparisons would indicate that the cumulative incidence of both AIDS and death continued to decline following the widespread availability of HAART since 1996. However, this cannot take account of changes in the population at risk or changes in the levels of risk due to the effect of HAART. Descriptive analysis is unable to show the extent to which changes in the frequency of events, such as AIDS diagnoses or deaths, are due to:

a) changes in CD4 cell counts;

b) changes in other risk factors for progression to AIDS or death due to changing characteristics of the population;

c) an actual reduction in the risk of AIDS or death independent of CD4 cell counts and other population characteristics.

The independent effects of CD4 cell counts and population characteristics were therefore investigated using multivariable Poisson regression analysis of the integrated dataset to compare incidence rates while adjusting for changes in other factors. The integrated dataset was therefore used to describe the

progression of infection among individuals to either AIDS or death as if they were being monitored as a prospective cohort. Such a 'pseudo-cohort' required definition of an individual's person-years of follow-up (PYFU) – the time during which reports of that individual were being received by surveillance systems indicating that events, such as AIDS or death, would be reported if they were to occur. This was a natural extension of the Poisson regression analysis of loss to follow-up (LTFU) to investigate potential bias in chapter 5. For this analysis, it was also important to be able to monitor each patient's changing level of risk over time to be able to determine the incidence rate of occurrence of each event per person per unit of time at each level of risk. This is because AIDS and death are highly associated with CD4 cell counts, which were used in this analysis as proxy markers of risk.

7.2 Aims

a) To use the integrated dataset to describe trends in the distribution of time spent in CD4 cell count categories by adults living with HIV in E,W&NI.

b) To use the integrated dataset to determine national trends in the incidence of first AIDS diagnoses and death, independent of other factors.

c) To use the integrated dataset to identify factors that were independently associated with the incidence of events and determine their relative effect.

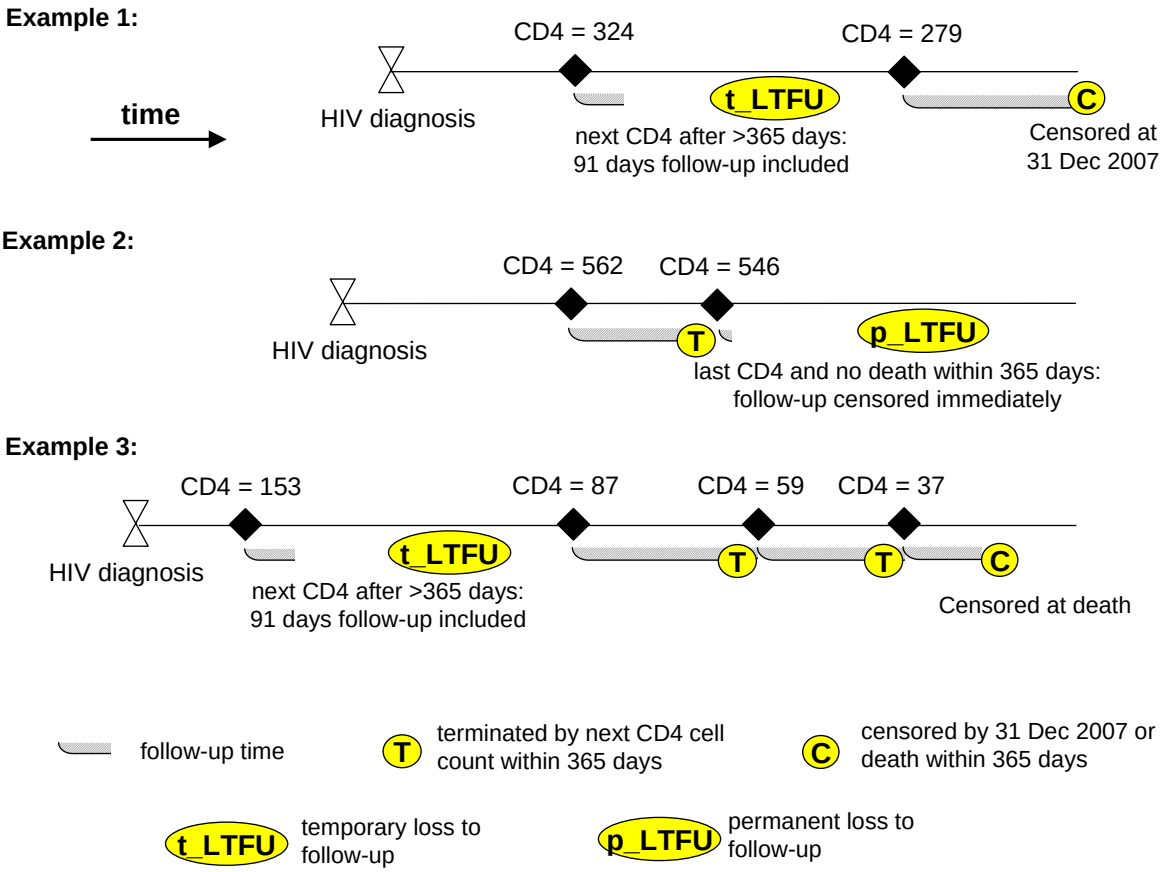
7.3 Methods

7.3.1 Data preparation

The methodology used to determine patient follow-up was similar to that described earlier in the analysis of LTFU (Section 5.3.3). CD4 cell counts in the integrated dataset were used to define the time over which each individual was followed in the cohort (to account for loss to follow-up, emigration and unreported deaths). Patient follow-up was split into distinct periods based on CD4 cell counts in the integrated dataset between 1995 and 2007 (Figure 7.1). Each period of follow-up was assumed to start on the date of the CD4 cell count and end at the earliest of a) the next CD4 cell count, b) a clinical event, c) 31st December 2007, d) 91 days following the date of each CD4 cell count if the CD4 cell count was followed by another but not within 365 days (temporary LTFU). Follow-up was censored immediately after the last CD4 cell count if it was before the 1st January 2007 or if it was more than 365 days before the date of death (permanent LTFU). Periods of follow-up defined by CD4 cell counts within 91 days of HIV diagnosis were considered to start at the date of diagnosis to account for the potential bias that could have been introduced by excluding follow-up and AIDS diagnoses during this period (Sections 5.4.4.3 and 5.5.5.2). Each period of follow-up, and any events that occurred in that period, were then allocated to a category of risk depending on the value of the CD4 count[448]. This allowed each individual's level of risk to change over time, but considered the level of risk to be constant (based on their CD4 count values) during each period of follow-up. Periods of follow-up were censored at a maximum of 91 days following the date of each CD4 cell count (subject to sensitivity analysis) during temporary LTFU because data gathered from

surveillance systems are likely to be incomplete and because imperfect record linkage within and between surveillance systems can lead to gaps in follow-up. Yet the patient can be considered to be under follow-up for part of this time, reflected in the 91 days, as most patients are only seen for clinical appointments three of four times a year. In contrast, periods of follow-up were censored immediately following the date of each CD4 cell count during permanent LTFU. If patient follow-up was censored by temporary LTFU, then they were allowed to re-enter the risk set if and when they had a further CD4 measurement. This ensured that PYFU were included in the analysis at the appropriate level of risk and that an event that occurred a long time after the last CD4 cell count was measured was not considered to occur at that level of risk. Therefore, only events that occurred within the defined PYFU were considered in the analysis. Because many individuals have three or four CD4 cell counts each year, time at risk was expected to be continuous for most individuals (Sections 5.4.3.2 and 5.4.3.3) although split into periods with time-updated values of exposures. In the analysis of the incidence of first AIDS diagnoses, any CD4 cell counts that occurred after the event were excluded from the analysis. Also, because only the month of AIDS is reported to HARS, PYFU were considered to end in AIDS when AIDS was reported in the same or following twelve months (subject to sensitivity analysis). PYFU were terminated halfway between the CD4 cell count and the end of the month if the CD4 count was measured in the same month as the AIDS event, and on the 15th of the month in which AIDS was diagnosed when AIDS was reported in one of the following twelve months.

Figure 7.1. Hypothetical calculation of periods of follow-up



PYFU and events were subdivided and allocated to calendar years in which they occurred. PYFU and events were also subdivided into those that occurred within the first six months after HIV diagnosis (subject to sensitivity analysis) and those that occurred subsequently. This allowed consideration of whether the risk of AIDS or death was higher in the first six months after HIV diagnosis than that which would have been expected based on the individuals' CD4 cell counts at the time. This was hypothesised to reflect induction into treatment and care and to contribute to the peak in deaths within the first six months after HIV diagnosis (Section 6.4.4).

As CD4 counts were not linearly associated with AIDS or death, CD4 categories were used to assign levels of risk for descriptive analyses but the square root transformation of the CD4 cell count was used for multivariable analysis (as recommended by the literature [Section 1.3.1] and shown by closer fitting to the trend in incidence than when using the logarithm of the CD4 cell count [Figure 7.4]). Age was included in the analysis as a continuous linear variable based on visual assessment of the plot of incidence by age (Figure 7.3) but was also presented in groups to simplify description of the data (Table 7.1).

7.3.2 Statistical analysis

Poisson regression was used for the analysis of incidence rates as described previously (Section 5.3.3).

The main effects model included the following:

1. time-updated continuous variables: square root of CD4 cell count; age
2. time-updated binary variables: previous AIDS diagnosis; region; PYFU within the first six months after HIV diagnosis
3. fixed categorical variables: calendar year of diagnosis; ethnicity; risk group.

To investigate the possibility of different trends in incidence within the first six months after HIV diagnosis and incidence during subsequent follow-up, the analysis was repeated including an interaction between calendar year and whether PYFU were within the first six months after HIV diagnosis.

7.3.3 Sensitivity analyses

Sensitivity analyses were carried out on the model for the incidence of death to determine the impact of varying the way in which PYFU were defined. Firstly, the analysis was repeated using the logarithm of the CD4 cell counts and categorisation of the CD4 cell counts. The analysis was also repeated, censoring the periods of follow-up as follows: 1) immediately or after 365 days during periods of temporary LTFU (deaths were still considered if within 365 days of the last CD4 cell count); 2) at a maximum of six months (182 days) following the date of each CD4 count (deaths were only considered if within 182 days of the last CD4 cell count) or only at the end of 2007 only – i.e. with no maximum; 3) at the end of continuous follow-up per individual – i.e. individuals were not allowed to re-enter the risk set and only the first continuous phase of follow-up was considered. The analysis was also repeated with: 4) termination of all follow-up at age 59 – to exclude the possibility of bias due to under-identification of deaths through record linkage to ONS death files (limited to individuals over 59 years of age); 5) Subdividing follow-up one year after HIV diagnosis instead of six months after HIV diagnosis. The following sensitivity analyses were carried out on the model for AIDS incidence: 1) consideration of only AIDS diagnoses reported to HARS; 2) termination of all follow-up at the end of 1999 – to exclude the period during which AIDS reporting may have significantly dropped; 3) exclusion of time between HIV diagnosis and first CD4 cell count for individuals with first CD4 cell count within 91 days of diagnosis.

7.4 Results: incidence of death (PYFU censored at 365 days)

7.4.1 Deaths

In the integrated dataset, there were 3,779 deaths with CD4 cell counts in the preceding year, increasing from 360 in 1995 to 555 in 1996, after which the numbers remained low (variation from 202 to 317). Deaths within the first six months of HIV diagnosis accounted for 14.0% (528/3,779) of these deaths increasing from 5.3% (19/360) in 1995 to 22.8% (72/316) in 2005 before falling to 17.6% (42/239) in 2007.

7.4.2 Person-years of follow-up (PYFU)

There were 830,904 CD4 cell counts from 49,677 individuals that defined PYFU in the risk set between 1995 and 2007 (Table 7.1). There was a median 3.67 years between the start and end of follow-up (IQR 1.36, 7.00, range 1 day to 13.1 years). A quarter (24.7% [12,278/49,677]) of individuals had gaps in their follow-up (summing up to a median 1.55 PYFU [IQR 0.93, 2.71] and adding five or more PYFU for 865 [1.7%] patients) so that, overall, individuals contributed a median 3.16 PYFU (IQR 1.15, 6.17, range 1 day to 13.1 years). There were a total of 204,233 PYFU, increasing from 3,929 in 1995 to 33,058 in 2007.

By calendar year, an increasing proportion of the total PYFU were spent with high CD4 cell counts (Figure 7.2a). PYFU with CD4 cell counts below 100 cells/mm^3 accounted for 26.0%, 11.0%, 6.1% and 2.7% of the total PYFU in 1995, 1998, 2001 and 2007 respectively. The equivalent trend for PYFU with CD4 cell counts of 350 cells/mm^3 or above was 33%, 43%, 55% and 68%. Absolute PYFU with CD4 counts below 100 cells/mm^3 remained substantial after 1995; 894, 784 and 892 years in 1998, 2001 and 2007 (Figure 7.2b).

Figure 7.2a. Proportional distribution of PYFU by CD4 count category over time

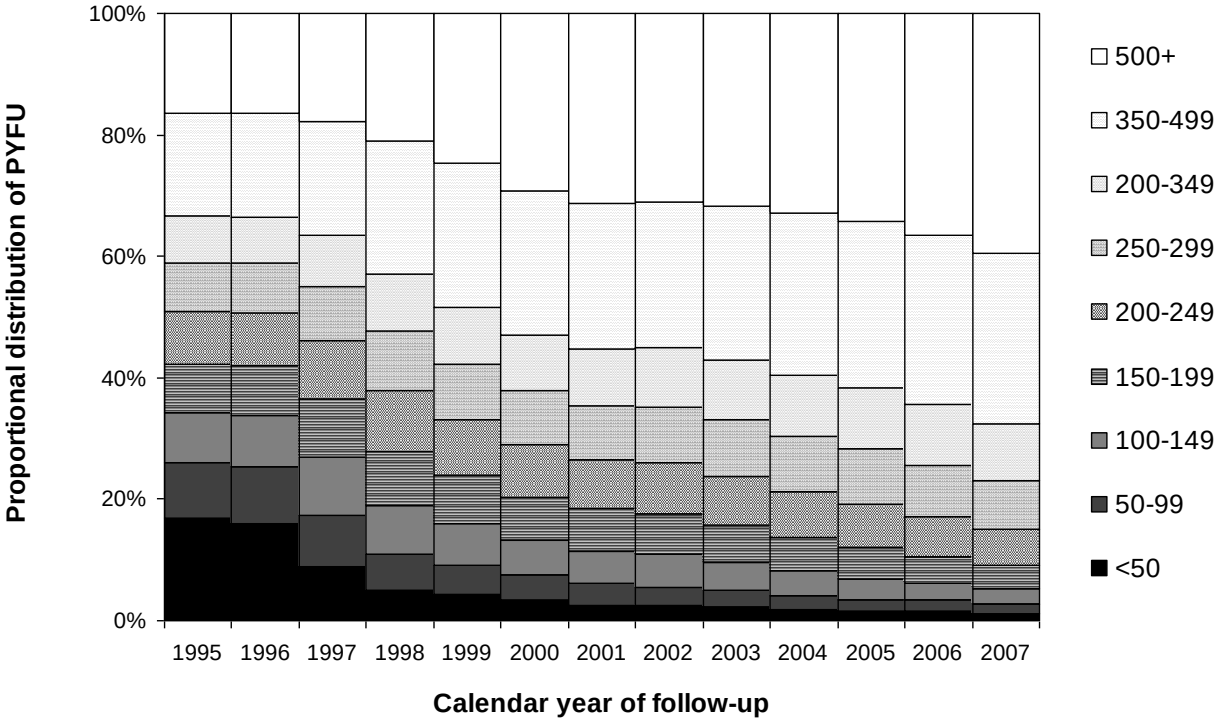
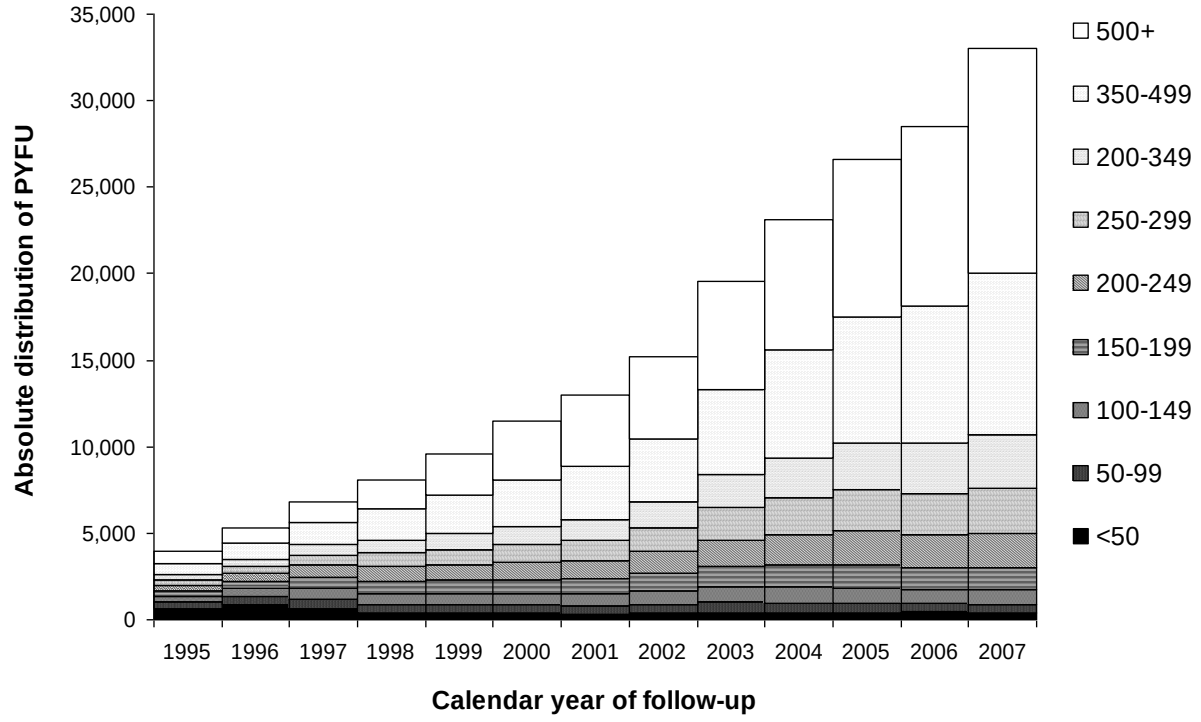


Figure 7.2b. Absolute distribution of PYFU by CD4 count category over time



PYFU within six months of HIV diagnosis accounted for 8.0% (16,415/204,233) of all PYFU and this proportion was fairly stable over time. PYFU within six months of HIV diagnosis accounted for an increasing proportion of all PYFU at CD4 cell counts less than 100 cells/mm^3 (increasing from 17.1% in 1995 to 61.7% in 2005 before falling to 50.6% in 2007) and a decreasing proportion of all PYFU at CD4 cell counts or 349 cells/mm^3 or more (decreasing from 26.8% in 1995 to 8.9% in 2007).

7.4.3 Incidence of death

There was an overall incidence of death of 18.5 (95% CI 17.9, 19.1) per 1,000 PYFU (Table 7.1). The incidence of death among MSM (19.4 [18.6, 20.2]) and heterosexual males (20.2 [18.7, 21.9]) was higher than that among heterosexual females (12.1 [11.1, 13.2]) and women diagnosed antenatally (3.3 [2.0, 5.0]) but lower than that among recipients of blood/blood products (36.1 [29.2, 44.2]) and IDUs (50.9 [45.4, 56.8]). Incidence among white (21.4 [20.6, 22.3]) and Indian/Pakistani/Bangladeshi individuals (21.5 [16.2, 28.1]) was higher than among other groups (varying from 13.2 [12.3, 14.2] among black Africans to 15.6 [12.9, 18.7] among individuals of other/mixed ethnicity). The incidence rate increased linearly with age (Figure 7.3). The incidence of death within the first six months after HIV diagnosis was 32.2 (29.5, 35.0) per 1,000 PYFU compared to 17.3 (16.7, 17.9) per 1,000 PYFU for those followed six months or more after HIV diagnosis. Incidence rates by CD4 cell count category were not linearly associated with death. The association between the square root of CD4 cell counts and incidence rates more closely approximated the incidence rates in each CD4 count category than the association with the logarithm of the CD4 count (Figure 7.4).

Figure 7.3. Incidence rates of death by age

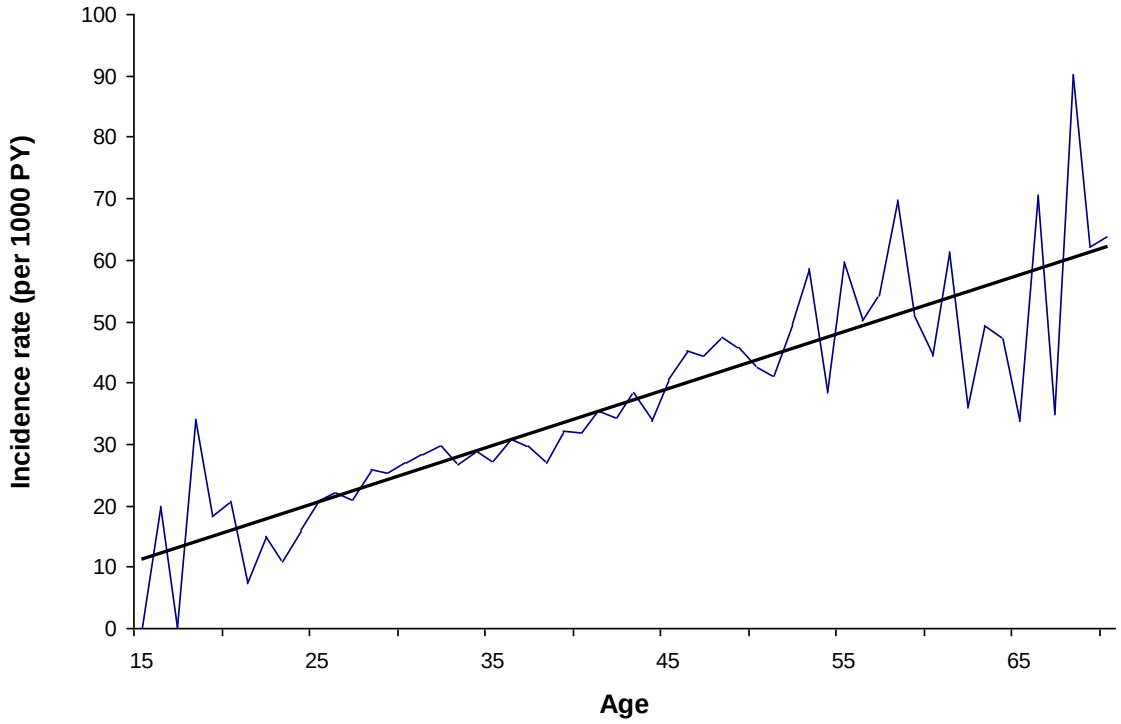
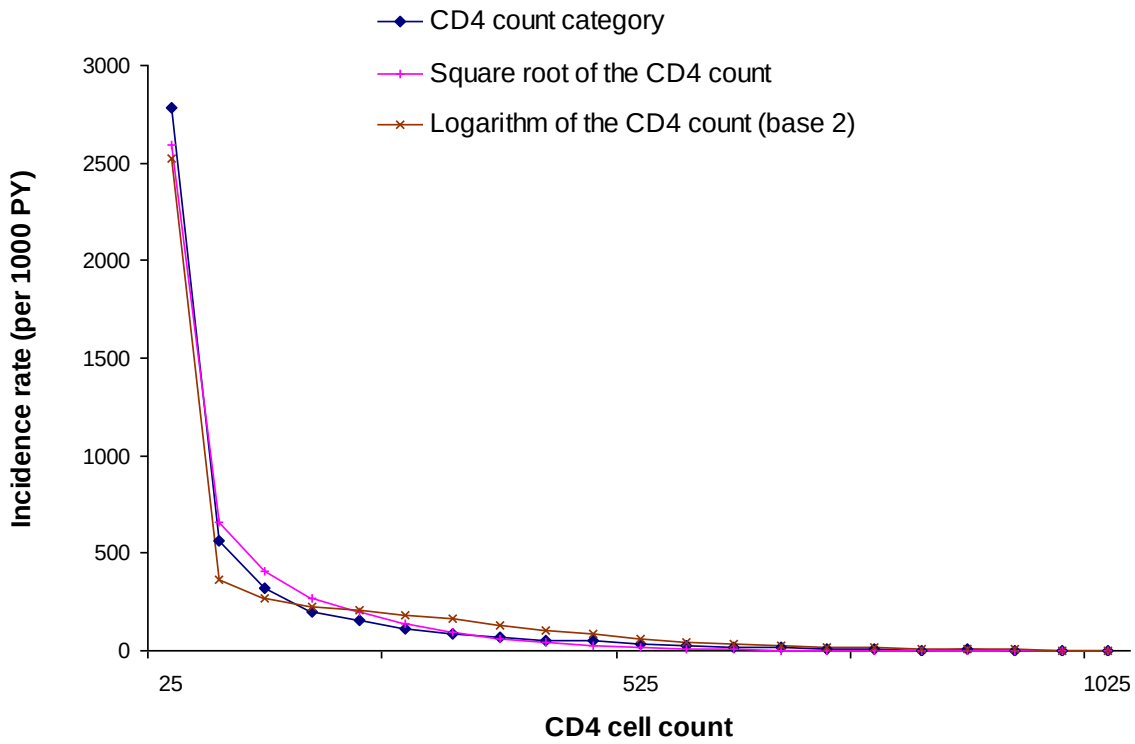


Figure 7.4. Modelled incidence rates of death by transformed CD4 cell counts



7.4.4 Trends in the incidence of death

The incidence of death increased from 91.6 (82.4, 101.6) per 1,000 PYFU in 1995 to 104.8 (96.3, 113.9) per 1,000 PYFU in 1996 before falling to 21.9 (19.3, 24.8) per 1,000 PYFU in 2000 and to 7.2 (6.3, 8.2) per 1,000 PYFU in 2007 (Table 7.1 and Figure 7.5). This overall trend over time was reflected in each CD4 cell count category (Figure 7.5). Trends in the incidence of death over time within the first six months after HIV diagnosis were less marked than those seen among deaths occurring more than six months after HIV diagnosis: 69.2 (47.6, 97.2) per 1,000 PYFU in 1996 to 43.9 (31.5, 59.6) in 2000 and to 22.2 (16.0, 30.0) in 2007 (Figure 7.6).

Figure 7.5. Trends in the incidence of death stratified by CD4 count category

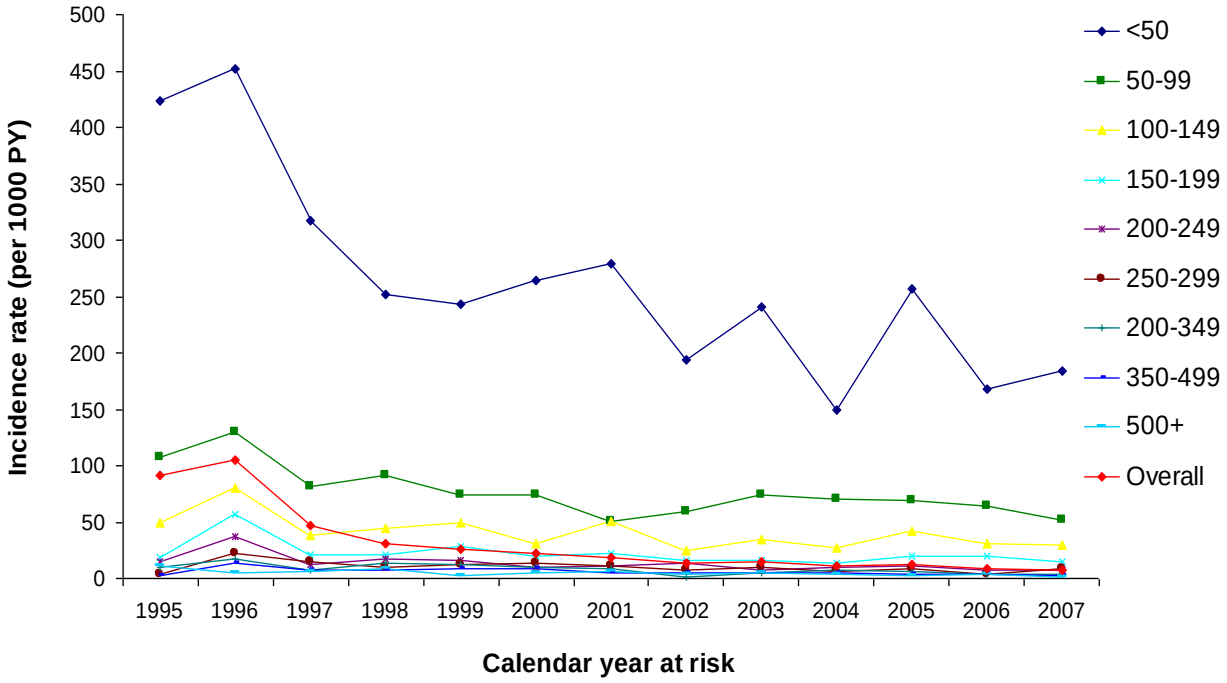


Figure 7.6. Trends in the incidence of death stratified according to follow-up within or more than six months after HIV diagnosis

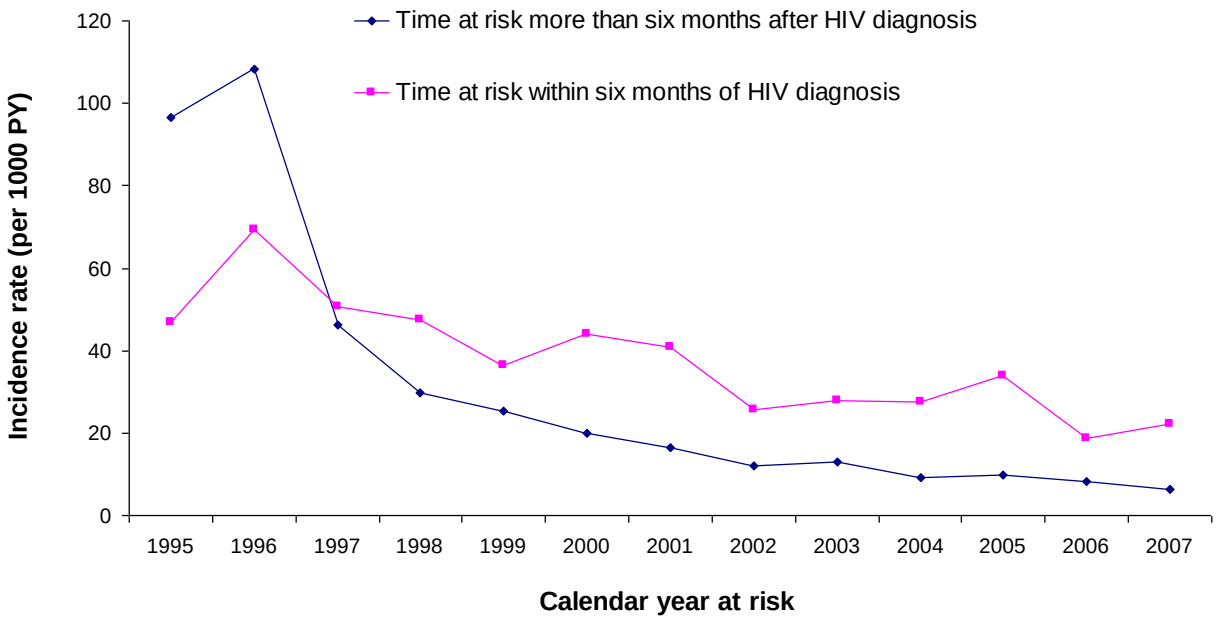


Table 7.1. Number of CD4 cell counts, PYFU and incidence of death stratified by demographic and other factors.

		Number of CD4 cell counts n (%)	Person-years of follow-up (%)	Number of deaths n (%)	Incidence rate (per 1,000 PYFU)
Calendar year					
	1995	19,682 (2)	3,929 (2)	360 (10)	91.6
	1996	23,375 (3)	5,294 (3)	555 (15)	104.8
	1997	29,719 (4)	6,820 (3)	317 (8)	46.5
	1998	33,699 (4)	8,103 (4)	252 (7)	31.1
	1999	41,874 (5)	9,618 (5)	252 (7)	26.2
	2000	49,813 (6)	11,444 (6)	251 (7)	21.9
	2001	56,035 (7)	12,957 (6)	244 (6)	18.8
	2002	64,730 (8)	15,167 (7)	202 (5)	13.3
	2003	82,711 (10)	19,541 (10)	282 (7)	14.4
	2004	95,137 (11)	23,153 (11)	251 (7)	10.8
	2005	109,482 (13)	26,599 (13)	316 (8)	11.9
	2006	109,078 (13)	28,540 (14)	258 (7)	9.0
	2007	115,569 (14)	33,058 (16)	239 (6)	7.2
PYFU within first six months after HIV diagnosis					
	No	729,565 (88)	187,816 (92)	3251 (86)	17.3
	Yes	101,339 (12)	16,417 (8)	528 (14)	32.2

CD4 cell count category				
<50	32,819 (4)	6,109 (3)	1,741 (46)	285.0
50-99	34,159 (4)	6,431 (3)	490 (13)	76.2
100-149	46,000 (6)	9,258 (5)	356 (9)	38.5
150-199	58,308 (7)	11,950 (6)	241 (6)	20.2
200-249	71,261 (9)	15,531 (8)	177 (5)	11.4
250-299	78,296 (9)	18,070 (9)	159 (4)	8.8
300-349	79,208 (10)	19,536 (10)	122 (3)	6.2
350-499	198,378 (24)	52,054 (25)	247 (7)	4.7
>499	232,475 (28)	65,290 (32)	246 (7)	3.8
Previous AIDS diagnosis				
No	538,971 (65)	136,429 (67)	1,001 (26)	7.3
Yes	291,933 (35)	67,804 (33)	2,778 (74)	41.0
Region				
Outside London	307,729 (37)	73,835 (36)	1,538 (41)	20.8
London	523,175 (63)	130,398 (64)	2,241 (59)	17.2
Age group				
15-25	30,520 (4)	7,716 (4)	56 (1)	7.3
25-29	86,629 (10)	21,746 (11)	282 (7)	13.0
30-34	160,769 (19)	39,954 (20)	647 (17)	16.2
35-39	193,279 (23)	47,638 (23)	766 (20)	16.1
40-44	152,151 (18)	37,350 (18)	678 (18)	18.2
45-49	92,998 (11)	22,414 (11)	548 (15)	24.4
50-54	53,981 (6)	12,958 (6)	323 (9)	24.9
>54	60,577 (7)	14,454 (7)	479 (13)	33.1
Risk group				
MSM	457,677 (55)	111,436 (55)	2,159 (57)	19.4
Heterosexual men	124,548 (15)	30,969 (15)	626 (17)	20.2
Heterosexual women	188,618 (23)	46,603 (23)	565 (15)	12.1
Heterosexual women (not diagnosed antenatally)	25,805 (3)	6,450 (3)	21 (1)	3.3
IDU	23,527 (3)	6,172 (3)	314 (8)	50.9
Recipients of blood/blood products	10,729 (1)	2,601 (1)	94 (2)	36.1
Ethnicity				
White	512,215 (62)	125,016 (61)	2,682 (71)	21.5
Black-African	234,175 (28)	57,792 (28)	764 (20)	13.2
Black Caribbean	21,555 (3)	5,658 (3)	82 (2)	14.5
Black – Other	12,906 (2)	3,322 (2)	47 (1)	14.1
Indian/Pakistani/Bangladeshi	10,015 (1)	2,507 (1)	54 (1)	21.5
Other Asian	9,681 (1)	2,374 (1)	32 (1)	13.5
Other/mixed	30,357 (4)	7,561 (4)	118 (3)	15.6
Total	830,904 (100)	204,233 (100)	3,779 (100)	18.5

7.4.5 Poisson Regression

Given the large number of deaths observed, there were significant variations in the incidence of death in univariable analyses for each factor considered (Table 7.2). In particular, the incidence rate decreased by a fifth for every unit increase in the square root of the CD4 cell count (IRR 0.79; 95% CI 0.79, 0.80). The incidence rate was more than five times higher for individuals with PYFU who had a previous AIDS diagnosis (IRR 5.58; 95% CI 5.19, 6.00). In addition, the incidence rate among individuals followed in the first six months after HIV diagnosis was substantially higher than that among individuals followed more than six months after HIV diagnosis (IRR 1.86; 95% CI 1.69, 2.04). The incidence rate among individuals under follow-up in 2007 was less than a tenth of that among individuals with follow-up in 1996 (IRR 0.07; 95% CI 0.06, 0.08).

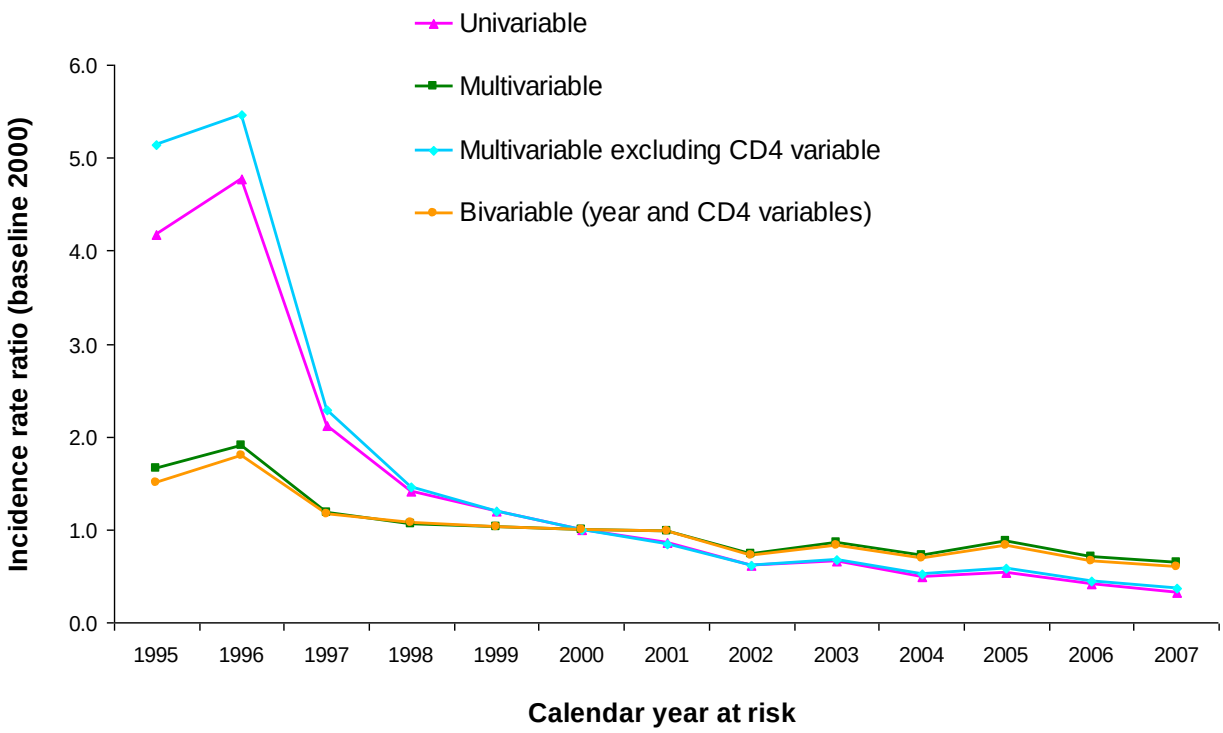
Adjustment for the CD4 cell count substantially decreased the decline in incidence observed between 1995 and 2007, particularly between 1995 and 1998 (bivariable analysis – Figure 7.7). Adjustment for factors other than the CD4 cell count in a multivariable analysis generally attenuated the incidence rate ratios for most factors studied except for age, PYFU within first six months after HIV diagnosis and IDU risk (Table 7.2). However, there was little difference between the bivariable analysis and multivariable analysis in the decline observed between 1995 and 2007. Adjustment for other factors except for CD4 cell count made no appreciable difference to the trend over time in comparison to the univariable model.

Table 7.2. Results from univariable and multivariable Poisson regression analysis of the association between each factor and the incidence of death

	Univariable incidence rate ratio (95% confidence interval)	Multivariable incidence rate ratio (95% confidence interval)
Calendar year		
1995	**4.18 (3.56, 4.91)**	**1.65 (1.40, 1.95)**
1996	**4.78 (4.12, 5.55)**	**1.91 (1.64, 2.22)**
1997	**2.12 (1.80, 2.5)**	1.18 (1.00, 1.39)
1998	**1.42 (1.19, 1.69)**	1.07 (0.89, 1.27)
1999	**1.19 (1.00, 1.42)**	1.03 (0.86, 1.23)
2000	-	-
2001	0.86 (0.72, 1.02)	0.99 (0.83, 1.18)
2002	**0.61 (0.50, 0.73)**	0.73 (0.61, 0.88)
2003	**0.66 (0.56, 0.78)**	0.87 (0.73, 1.03)
2004	**0.49 (0.41, 0.59)**	**0.73 (0.61, 0.87)**
2005	**0.54 (0.46, 0.64)**	0.87 (0.74, 1.03)
2006	**0.41 (0.35, 0.49)**	**0.70 (0.59, 0.84)**
2007	**0.33 (0.28, 0.39)**	**0.64 (0.54, 0.77)**
Square root of CD4 cell count	**0.79 (0.79, 0.80)**	**0.82 (0.82, 0.83)**
PYFU within first six months after HIV diagnosis		
No	-	-
Yes	**1.86 (1.69, 2.04)**	**1.35 (1.22, 1.49)**
Previous AIDS diagnosis		
Yes	**5.58 (5.19, 6.00)**	**2.34 (2.16, 2.53)**
No	-	-
Region		
Outside London	**1.21 (1.14, 1.29)**	**1.17 (1.10, 1.26)**
London	-	-
Age, per additional 10 years	**1.33 (1.29, 1.38)**	**1.37 (1.33, 1.42)**
Risk group		
MSM	-	-
Heterosexual men	1.04 (0.95, 1.14)	**0.87 (0.78, 0.97)**
Heterosexual women	**0.63 (0.57, 0.69)**	**0.83 (0.73, 0.94)**
Heterosexual women (diagnosed antenatally)	**0.17 (0.11, 0.26)**	**0.64 (0.41, 0.99)**
IDU	**2.63 (2.33, 2.96)**	**2.06 (1.82, 2.32)**
Recipients of blood products	**1.87 (1.52, 2.29)**	1.13 (0.92, 1.40)

Ethnicity		
White	-	-
Black African	**0.62 (0.57, 0.67)**	0.91 (0.81, 1.02)
Black Caribbean	**0.68 (0.54, 0.84)**	0.90 (0.72, 1.12)
Black Other	**0.66 (0.49, 0.88)**	0.87 (0.65, 1.17)
Indian\Pakistani\Bangladeshi	1.00 (0.77, 1.31)	0.97 (0.73, 1.27)
Other Asian	**0.63 (0.44, 0.89)**	0.83 (0.59, 1.18)
Other/mixed	**0.73 (0.60, 0.87)**	0.89 (0.74, 1.07)

Figure 7.7. Incidence rate ratio of death by calendar year of follow-up: comparison of different models



7.4.6 Interaction between calendar year and follow-up within six months of HIV diagnosis

The model including an interaction between calendar year and whether follow-up was within six months of HIV diagnosis or not indicated that incidence of death during follow-up within six months after HIV diagnosis did not decline over time (Table 7.3). In contrast, the rate ratio for follow-up more than six months after HIV diagnosis fell significantly between 1996 and 1997 (IRR 0.61, 95% CI 0.53, 0.70), but then decreased more gradually to 2007 (Table 7.3). The rate ratios for risk groups were attenuated (except that heterosexual men had a lower rate of death than MSM in this model) and the effects of ethnic groups became non-significant. The rate ratios for CD4 cell counts, age, previous AIDS diagnosis and region did not change significantly.

Table 7.3. Results from multivariable Poisson regression of the association between each factor and the incidence of death including an interaction between calendar year and PYFU within first six months after HIV diagnosis

		Multivariable incidence rate ratio (95% confidence interval)	**Multivariable incidence rate ratio: interaction between calendar year and PYFU within six months after HIV diagnosis (95% confidence interval)**
Square root of CD4 cell count		**0.82 (0.82, 0.83)**	**0.82 (0.82, 0.83)**
PYFU within first six months after HIV diagnosis			
	No	-	-
	Yes	**1.35 (1.22, 1.49)**	**1.54 (1.10, 2.15)**
Previous AIDS diagnosis			
	No	**-**	-
	Yes	**2.34 (2.16, 2.53)**	**2.35 (2.18, 2.54)**
Age, per additional 10 years		**1.37 (1.33, 1.42)**	**1.37 (1.33, 1.42)**

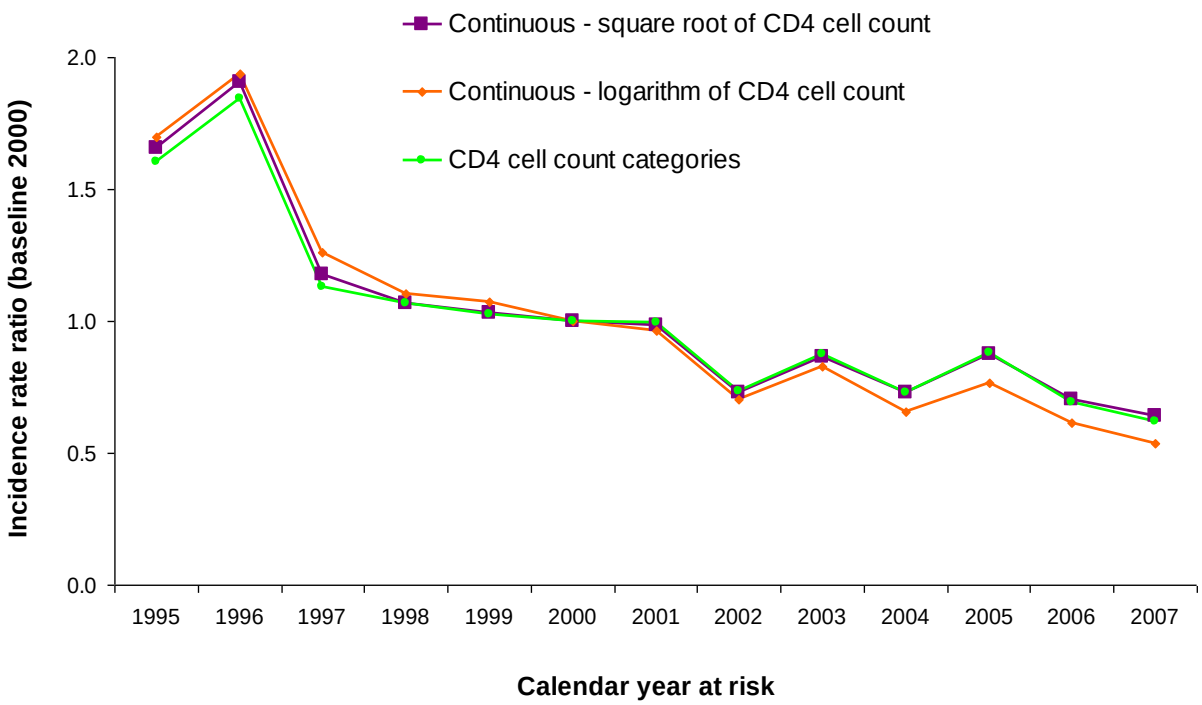
Calendar year		all PYFU	PYFU more than six months after HIV diagnosis
	1995	**4.18 (3.56, 4.91)**	**1.77 (1.48, 2.11)**
	1996	**4.78 (4.12, 5.55)**	**2.01 (1.71, 2.36)**
	1997	**2.12 (1.80, 2.50)**	**1.22 (1.02, 1.46)**
	1998	**1.42 (1.19, 1.69)**	1.10 (0.91, 1.33)
	1999	**1.19 (1.00, 1.42)**	1.08 (0.89, 1.30)
	2000	-	-
	2001	0.86 (0.72, 1.02)	0.97 (0.79, 1.17)
	2002	**0.61 (0.50, 0.73)**	**0.75 (0.61, 0.92)**
	2003	**0.66 (0.56, 0.78)**	0.88 (0.73, 1.06)
	2004	**0.49 (0.41, 0.59)**	**0.71 (0.58, 0.86)**
	2005	**0.54 (0.46, 0.64)**	0.84 (0.69, 1.01)
	2006	**0.41 (0.35, 0.49)**	**0.74 (0.61, 0.89)**
	2007	**0.33 (0.28, 0.39)**	**0.63 (0.51, 0.76)**
			PYFU within six months after HIV diagnosis
	1995		**0.50 (0.28, 0.89)**
	1996		0.62 (0.38, 1.01)
	1997		0.76 (0.46, 1.27)
	1998		0.81 (0.48, 1.35)
	1999		0.73 (0.43, 1.22)
	2000		-
	2001		1.09 (0.69, 1.71)
	2002		0.86 (0.53, 1.40)
	2003		0.90 (0.57, 1.40)
	2004		1.10 (0.70, 1.72)
	2005		1.20 (0.78, 1.84)
	2006		0.72 (0.44, 1.18)
	2007		1.16 (0.72, 1.86)
Region			
	Outside London	**1.21 (1.14, 1.29)**	**1.17 (1.10, 1.25)**
	London	-	-
Risk group			
	MSM	-	-
	Heterosexual men	1.04 (0.95, 1.14)	**0.87 (0.78, 0.97)**
	Heterosexual women	**0.63 (0.57, 0.69)**	**0.83 (0.73, 0.94)**
	Heterosexual women (diagnosed antenatally)	**0.17 (0.11, 0.26)**	**0.63 (0.41, 0.98)**
	IDU	**2.63 (2.33, 2.96)**	**2.07 (1.83, 2.33)**
	Recipients of blood products	**1.87 (1.52, 2.29)**	1.13 (0.91, 1.39)
Ethnicity			
	White	-	-
	Black African	**0.62 (0.57, 0.67)**	0.90 (0.80, 1.02)
	Black Caribbean	**0.68 (0.54, 0.84)**	0.90 (0.72, 1.13)
	Black Other	**0.66 (0.49, 0.88)**	0.88 (0.65, 1.17)
	Indian\Pakistani\Bangladeshi	1.00 (0.77, 1.31)	0.97 (0.73, 1.27)
	Other Asian	**0.63 (0.44, 0.89)**	0.82 (0.58, 1.17)
	Other/mixed	**0.73 (0.60, 0.87)**	0.88 (0.73, 1.06)

7.4.7 Sensitivity analysis

7.4.7.1 Transformation/categorisation of CD4 cell counts

Replacement of the square root of the CD4 cell count in the multivariable analysis with CD4 categories had little effect on the incidence rate ratios or the trends observed (Figure 7.8). Use of the logarithm of the CD4 cell count increased the rate ratios before 2000 and reduced them afterwards such that there was a slightly more marked trend over time.

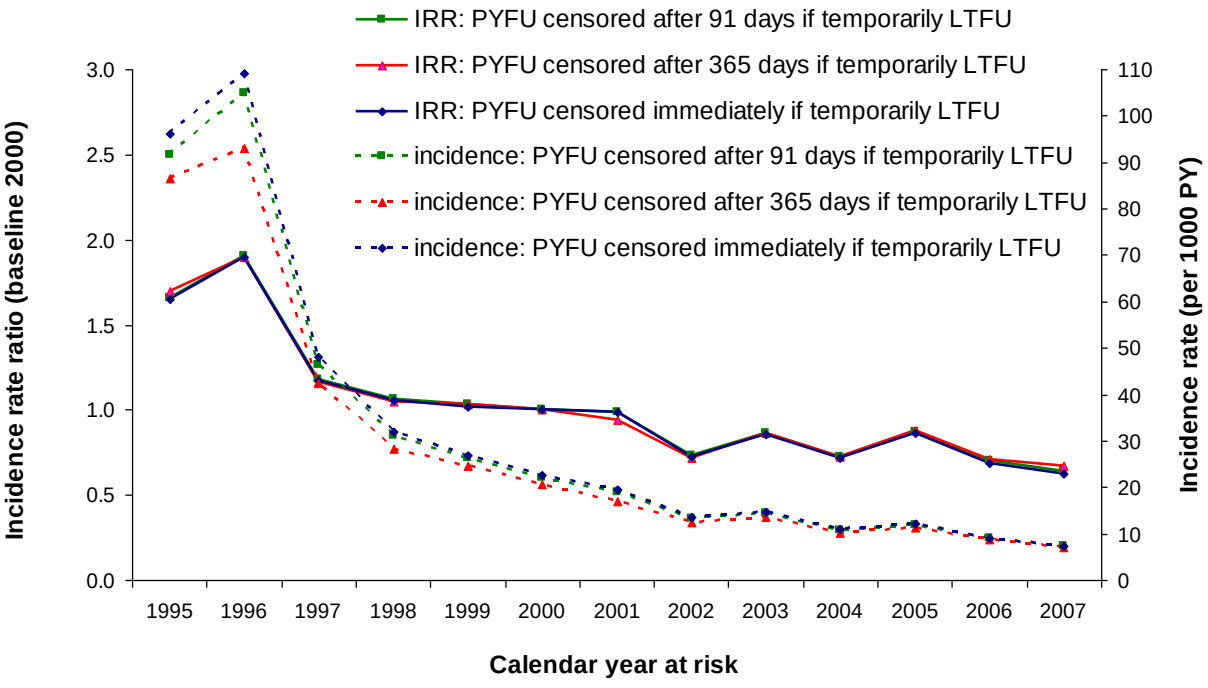
Figure 7.8. Incidence rate ratio of death by calendar year of follow-up: multivariable models using CD4 count categories and transformations of the CD4 count compared to univariable analysis



7.4.7.2 Changing the strategy for censoring PYFU during temporary follow-up

The number of PYFU increased to 215,449 and incidence decreased to 17.5 (95% CI 17.0, 18.1) per 1,000 PYFU when the analysis was repeated with inclusion of all 365 days of follow-up time during periods of temporary LTFU. When follow-up was censored immediately during temporary LTFU, the number of PYFU decreased to 200,550 and incidence increased to 18.8 (95% CI 18.3, 19.5) per 1,000 PYFU. The multivariable incidence rate ratios and estimates of incidence between 1995 and 2007 were almost exactly the same between the three models except that incidence estimates were reduced in 1995 and 1996 with less follow-up included (Figure 7.9). There were no major differences between the models in the rate ratios for the other determining factors.

Figure 7.9. Incidence rates and rate ratios of death by calendar year of follow-up – comparison of multivariable models using different censoring strategies with temporary LTFU defined as no subsequent measurement within 365 days of a CD4 cell count.

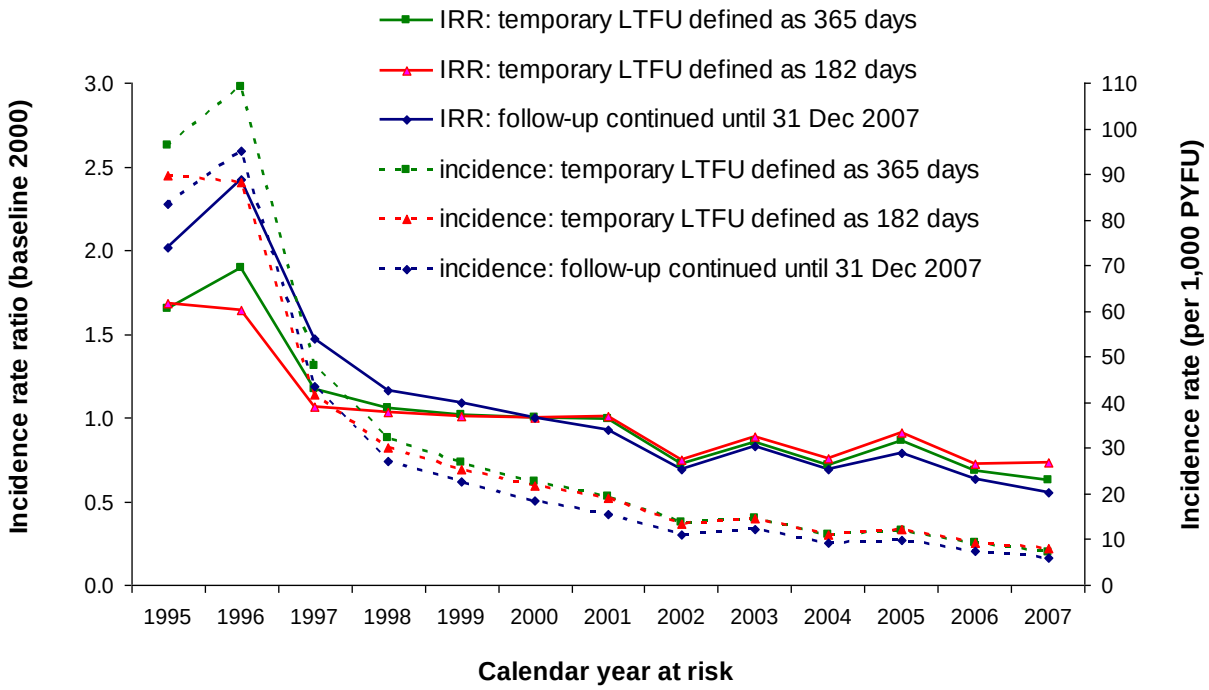


7.4.7.3 Changing the definition of temporary LTFU

When the definition of temporary LTFU was changed from 365 days, with follow-up time and deaths included for up to 365 days after a CD4 cell count measurement, to 182 days, with follow-up time and deaths included for up to 182 days, overall estimates of incidence increased marginally from 17.5 (95% CI 17.0, 18.1) to 17.9 (95% CI 17.3, 18.6) (total PYFU decreased from 215,449 to 180,821 years) (Figure 7.9). Without any temporary LTFU (follow-up after a CD4 cell count measurement considered to continue until 31 December 2007 or death), overall estimates of incidence declined to 15.6 (95% CI 15.1, 16.1) (total PYFU increased to 268,076 years). The number of deaths and PYFU included in the analysis increased with the maximum period of follow-up used for the definition of temporary LTFU and the number of individuals with gaps in their overall follow-up of more than five years declined from 1,472 to 865 to zero.

When temporary LTFU was defined by more than 182 days with no subsequent CD4 cell count, estimates of incidence were lower between 1995 and 1998 but there was little difference between the estimates for 1999 to 2007 (Figure 7.10). Rate ratios were also slightly lower in 1996 and 1997. With follow-up continued until 31 December 2007 or death, estimates of incidence were lower throughout than with temporary LTFU defined by more than 365 days with no subsequent CD4 cell count, but particularly in 1995 and 1996. Rate ratios were relatively high before 2000 and low after 2000 (Figure 7.10). However, overall trends for both incidence and the rate ratios were similar between the three models with different definitions of temporary LTFU. Differences in the rate ratios for the other factors were negligible between the models with defined temporary LTFU but slightly more marked for the model without temporary LTFU.

Figure 7.10. Incidence rates and rate ratios of death by calendar year of follow-up – comparison of multivariable models using different definitions of temporary LTFU.



7.4.7.4 Only considering the first continuous phase of follow-up

The total number of PYFU (158,961 years) and deaths (2,810) included in the model decreased when only the first continuous phase of follow-up (temporary LTFU defined by more than 365 days with no subsequent CD4 cell count measurement) was considered. However, estimated overall incidence, at 17.7 (95% CI 17.0, 18.3) per 1,000 PYFU, was little different than for other models. The incidence estimates and multivariable incidence rates ratios did not differ markedly over time between the models with continuous follow-up and those with gaps in the follow-up.

7.4.7.5 Excluding follow-up after 59 years of age

Exclusion of follow-up after 59 years of age reduced the overall PYFU to 197,116 and the number of deaths to 3,533 resulting in an overall incidence of 17.9 (95% CI 17.3, 18.5) per 1,000 PYFU. There were no substantial changes in incidence rate ratios in the model when compared to the original model including individuals of all ages.

7.4.7.6 Follow-up split one year after HIV diagnosis instead of six months after HIV diagnosis

When PYFU and events were subdivided into those that occurred within the first year after HIV diagnosis and those that occurred subsequently, the total number of PYFU and deaths included in the analysis did not change. The adjusted incidence rate ratio for follow-up within the first year after HIV diagnosis compared to later follow-up was 1.21 (95% CI 1.11, 1.32). This was slightly lower than in the original analysis (IRR 1.35; 95% CI 1.22, 1.49) and in the other sensitivity analyses. The rate ratios for the other factors did not differ substantially in this model compared to the original.

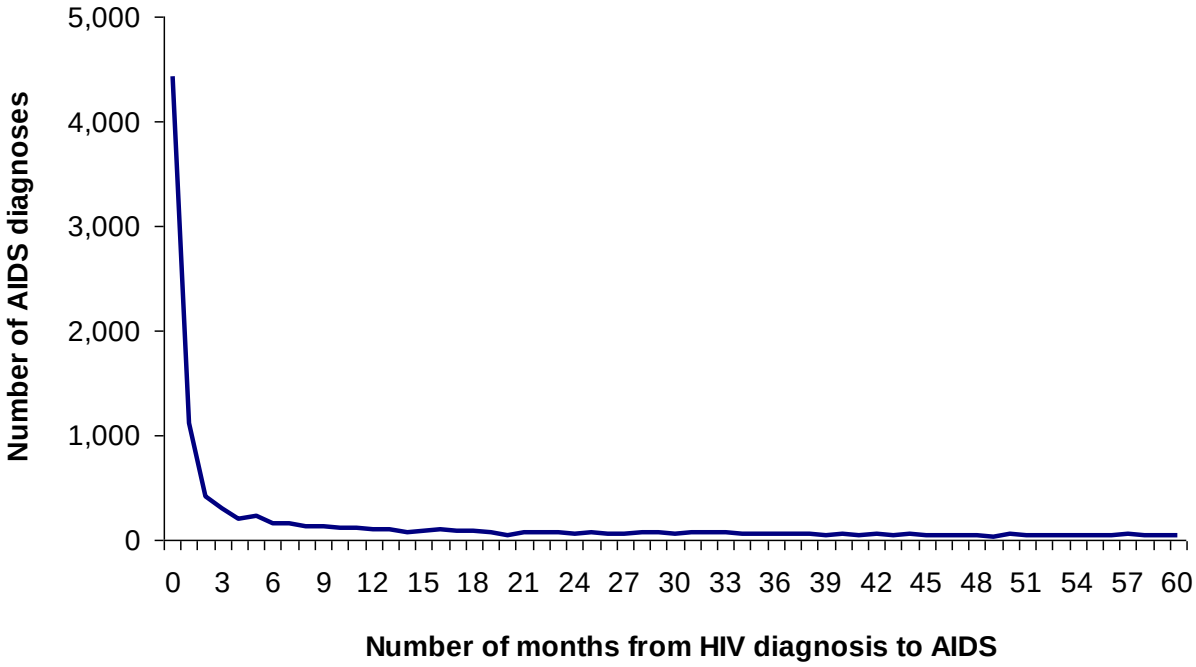
7.5 Results: incidence of AIDS (PYFU censored at 365 days)

7.5.1 AIDS

Over a quarter of individuals in the integrated dataset (31.5% [14,647/46,527]) were reported to have had AIDS. Nine in ten (93.4% [13,675]) of these AIDS diagnoses followed a CD4 cell count in 1994 or later and 97.4% (13,323) of those were less than 12 months after a CD4 cell count and were considered in this analysis. Half (50.7% [6,760]) of these AIDS diagnoses were reported to HARS (may also have been reported to SOPHID) and half were reported to

SOPHID but not HARS (49.3% [6,563]). There was a high number of AIDS diagnoses within the first three months after HIV diagnosis after which numbers of AIDS diagnoses were low but remained slightly elevated during the first year (Figure 7.11).

Figure 7.11. Number of AIDS diagnoses among all individuals who were known to have AIDS by the time from HIV diagnosis to AIDS (AIDS diagnoses within five years of HIV diagnosis and 12 months of a CD4 cell count)



The number of AIDS diagnoses per calendar year varied from 910 to 1,089 between 1995 and 1997 then fell to 698 in 1998. After 1998, AIDS diagnoses increased to 1,336 in 2003 then declined to 931 in 2007. AIDS within the first six months of HIV diagnosis occurred much more frequently than death within the first six months of HIV diagnosis. AIDS within the first six months of HIV diagnosis accounted for half (52.7% [7,015/13,323]) of the total number of AIDS

diagnoses. This percentage increased from 26.8% (239/891) in 1995 to 65.4% (753/1,152) in 2002 before declining to 51.9% (483/931) in 2007.

The percentage of AIDS diagnoses that were reported to HARS declined from 83.8% (747/891) in 1995 to 36.1% (336/931) in 2007, primarily due to a marked decline in HARS reports of AIDS diagnosed more than six months after HIV diagnosis (Figure 7.12). Therefore, the trend in the percentage of AIDS diagnoses that were reported to HARS was more marked for AIDS diagnoses that occurred more than six months after an HIV diagnosis (Figure 7.13).

Figure 7.12. Numbers of AIDS diagnoses reported to HARS and SOPHID by whether AIDS was within the first six months after an HIV diagnosis or not

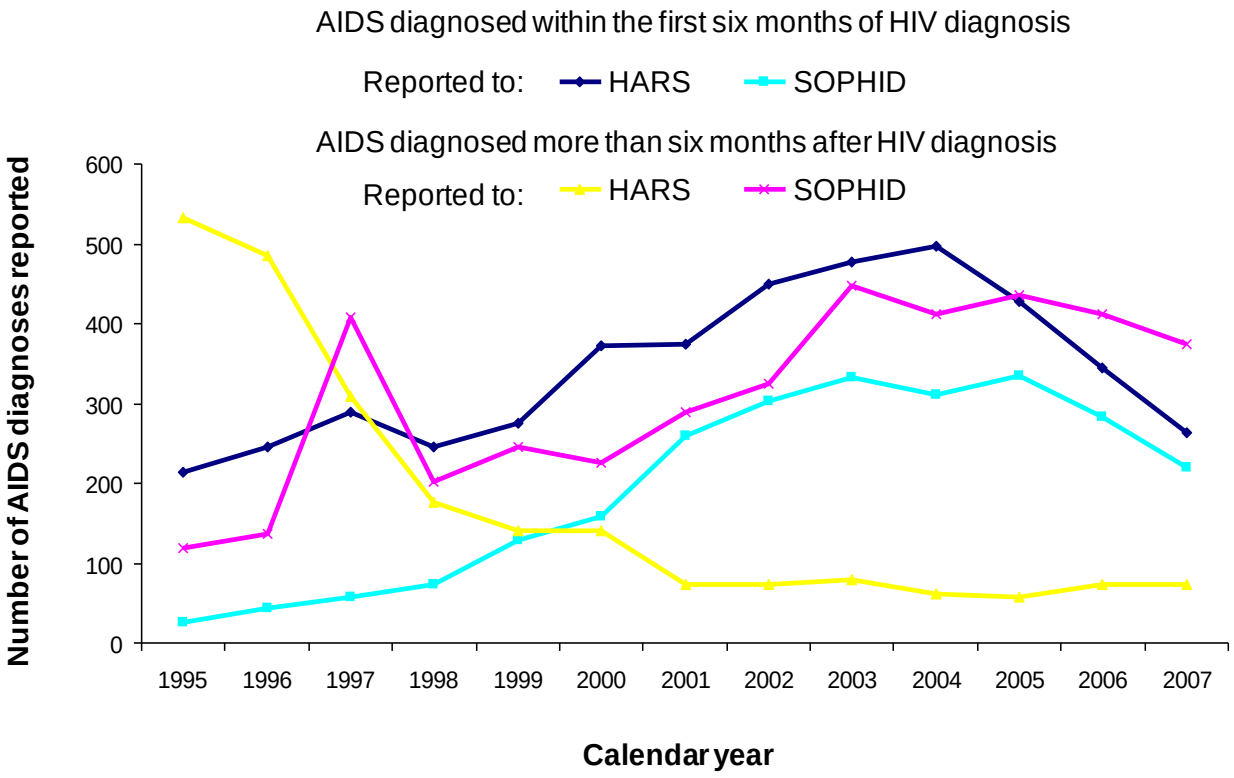
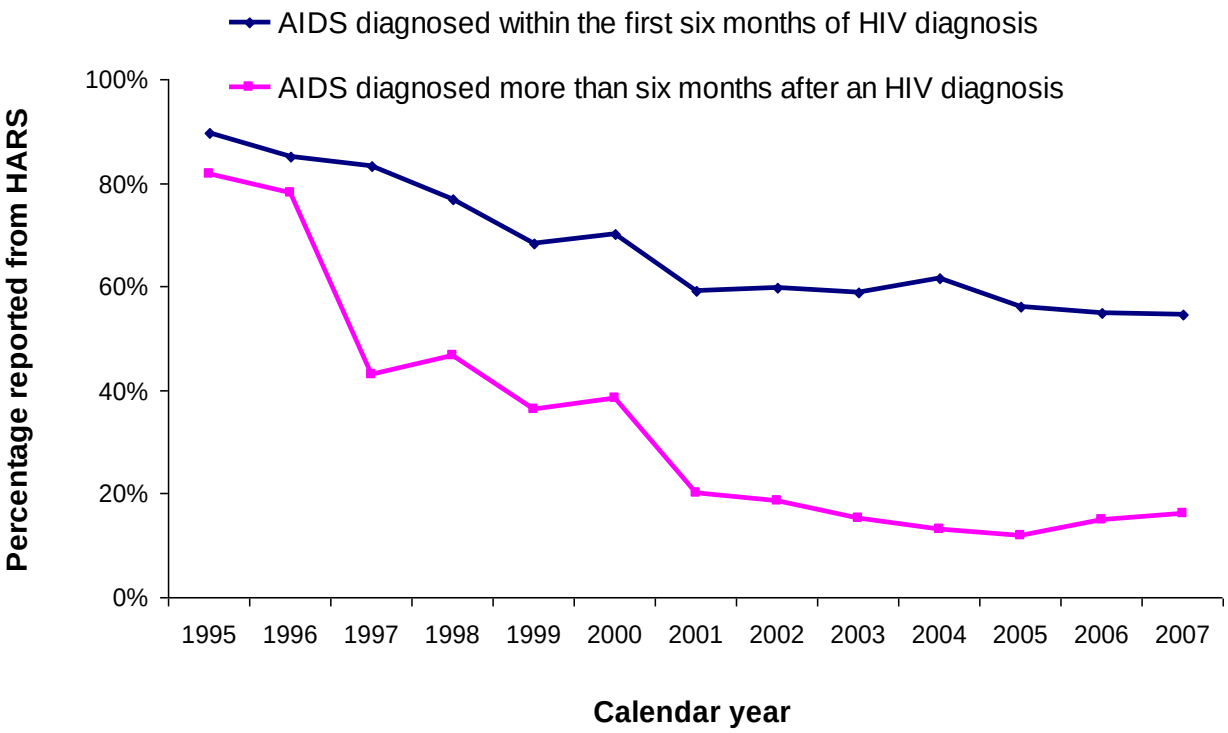


Figure 7.13. Percentage of AIDS diagnoses reported from HARS by whether AIDS was diagnosed within the first six months after an HIV diagnosis or not



7.5.2 Person-years of follow-up

There were 547,848 CD4 cell counts from 45,176 individuals that defined PYFU in the risk set between 1995 and 2007 (Table 7.6). Follow-up (median 2.0 PYFU [IQR 0.5, 4.6, range 1 day to 13.1 years] and the median number of CD4 cell counts during follow-up (8 [IQR 3, 18, range 1 to 124]) were slightly less per individual than for the analysis of incidence of death because PYFU between AIDS and death was excluded. There were a total of 136,648 PYFU, increasing from 3,436 in 1995 to 23,060 in 2007.

The distribution of PYFU by CD4 cell count was very similar but lower than that in the analysis of death incidence. PYFU with CD4 cell counts below 100 cells/mm^3 accounted for 12.9%, 5.3%, 3.4% and 1.7% of the total PYFU in

1995, 1998, 2001 and 2007 respectively. The equivalent trend for PYFU with CD4 cell counts of 350 cells/mm^3 or above was 43.9%, 54.3%, 61.5% and 69.8%. Absolute PYFU with CD4 cell counts below 100 cells/mm^3 declined from 441.9 in 1995 to 264.6 in 1998 then increased to 402.2 in 2003 and then remained stable at a mean of 394.0 PYFU between 2003 and 2007.

7.5.3 Incidence of AIDS

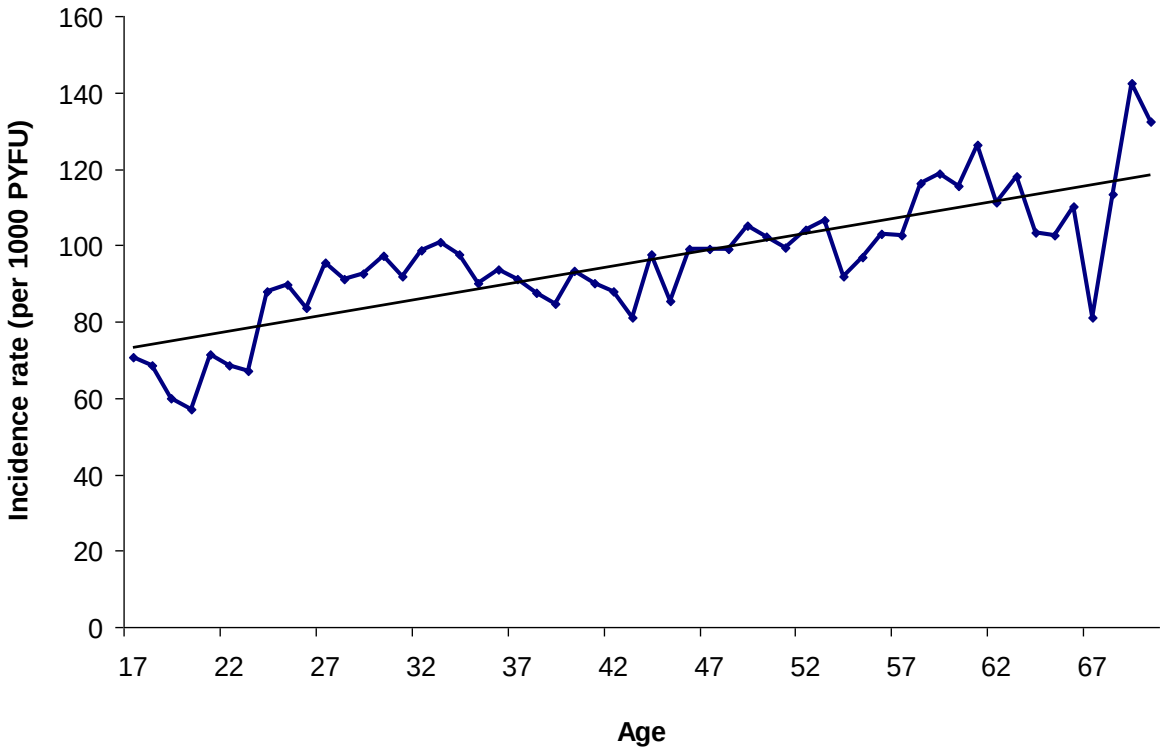
There was an overall incidence of AIDS of 93.0 (95% CI 91.4, 94.6) per 1,000 PYFU (Table 7.4). The incidence of AIDS within the first six months after HIV diagnosis was 469.0 (458.2, 479.9) per 1,000 PYFU compared to 44.9 (43.7, 46.1) per 1,000 PYFU for those followed six months or more after HIV diagnosis. The incidence of AIDS was highest among heterosexual men (152.5 [147.0, 158.2]), then in order of decreasing incidence, IDU (119.4 [108.7, 130.8]), heterosexual women (112.6 [108.9, 116.4]), recipients of blood/blood products (83.9 [71.0, 98.5]), MSM (73.1 [71.2, 75.1]) and finally, women diagnosed antenatally (35.2 [30.5, 40.4]). Incidence was highest among Indian/Pakistani/Bangladeshi individuals (145.0 [126.2, 165.9]), similar among other Asians (134.4 [116.3, 154.5]) and black Africans (127.0 [123.4, 130.6]) and lowest among white individuals (76.5 [74.6, 78.4]). The increase in the incidence rate was approximately linear with age (Figure 7.14).

Table 7.4. Number of CD4 cell counts, PYFU and incidence of AIDS stratified by demographic and other factors.

		Number of CD4 cell counts n (%)	Person-years of follow-up (%)	Number of AIDS diagnoses n (%)	Incidence rate (per 1,000 PYFU)
Calendar year					
	1995	14,014 (3)	3,435 (3)	978 (8)	284.6
	1996	14,974 (3)	3,589 (3)	819 (6)	228.2
	1997	18,280 (3)	4,425 (3)	973 (8)	219.9
	1998	20,172 (4)	5,018 (4)	647 (5)	128.9
	1999	24,850 (5)	5,933 (4)	734 (6)	123.7
	2000	30,079 (5)	7,089 (5)	851 (7)	120.0
	2001	34,953 (6)	8,178 (6)	934 (7)	114.2
	2002	41,723 (8)	9,776 (7)	1093 (9)	111.8
	2003	54,647 (10)	12,910 (9)	1,282 (10)	99.3
	2004	63,850 (12)	15,531 (11)	1,224 (10)	78.8
	2005	74,556 (14)	18,145 (13)	1,211 (10)	66.7
	2006	74,705 (14)	19,554 (14)	1,075 (8)	55.0
	2007	81,045 (15)	23,059 (17)	885 (7)	38.4
PYFU within first six months after HIV diagnosis					
	No	460,388 (84)	121,152 (89)	5,439 (43)	44.9
	Yes	87,460 (16)	15,495 (11)	7,267 (57)	469.0
CD4 cell count category					
	<50	12,845 (2)	1,884 (1)	3,894 (31)	2066.4
	50-99	15,328 (3)	2,695 (2)	1,981 (16)	734.9
	100-149	23,653 (4)	4,610 (3)	1,369 (11)	296.9
	150-199	33,401 (6)	6,618 (5)	1,050 (8)	158.7
	200-249	45,068 (8)	9,651 (7)	959 (8)	99.4
	250-299	52,814 (10)	12,094 (9)	723 (6)	59.8
	300-349	55,378 (10)	13,716 (10)	618 (5)	45.1
	350-499	142,139 (26)	37,687 (28)	1,155 (9)	30.6
	>499	167,222 (31)	47,688 (35)	957 (8)	20.1
Region					
	Outside London	209,412 (38)	50,992 (37)	5,072 (40)	99.5
	London	338,436 (62)	85,655 (63)	7,634 (60)	89.1
Age group					
	15-25	27,086 (5)	6,898 (5)	504 (4)	73.1
	25-29	70,199 (13)	17,804 (13)	1,620 (13)	91.0
	30-34	116,316 (21)	29,357 (21)	2,861 (23)	97.5
	35-39	126,719 (23)	31,754 (23)	2,843 (22)	89.5
	40-44	92,416 (17)	22,963 (17)	2,065 (16)	89.9
	45-49	53,219 (10)	12,947 (9)	1,248 (10)	96.4
	50-54	29,707 (5)	7,179 (5)	725 (6)	101.0
	>54	32,186 (6)	7,741 (6)	840 (7)	108.5

Risk group				
MSM	300,490 (55)	74,949 (55)	5,478 (43)	73.1
Heterosexual men	75,578 (14)	18,832 (14)	2,872 (23)	152.5
Heterosexual women	126,874 (23)	31,430 (23)	3,540 (28)	112.6
Heterosexual women (not diagnosed antenatally)	23,251 (4)	5,769 (4)	203 (2)	35.2
IDU	14,633 (3)	3,878 (3)	463 (4)	119.4
Recipients of blood/blood products	7,022 (1)	1,787 (1)	150 (1)	83.9
Ethnicity				
White	336,200 (61)	84,016 (61)	6,428 (51)	76.5
Black-African	155,617 (28)	38,249 (28)	4,857 (38)	127.0
Black Caribbean	15,077 (3)	3,975 (3)	389 (3)	97.8
Black – Other	8,780 (2)	2,302 (2)	212 (2)	92.1
Indian/Pakistani/Bangladeshi	5,768 (1)	1,461 (1)	212 (2)	145.0
Other Asian	6,025 (1)	1,466 (1)	197 (2)	134.4
Other/mixed	20,381 (4)	5,175 (4)	411 (3)	79.4
Total	547,848 (100)	136,647 (100)	12,706 (100)	93.0

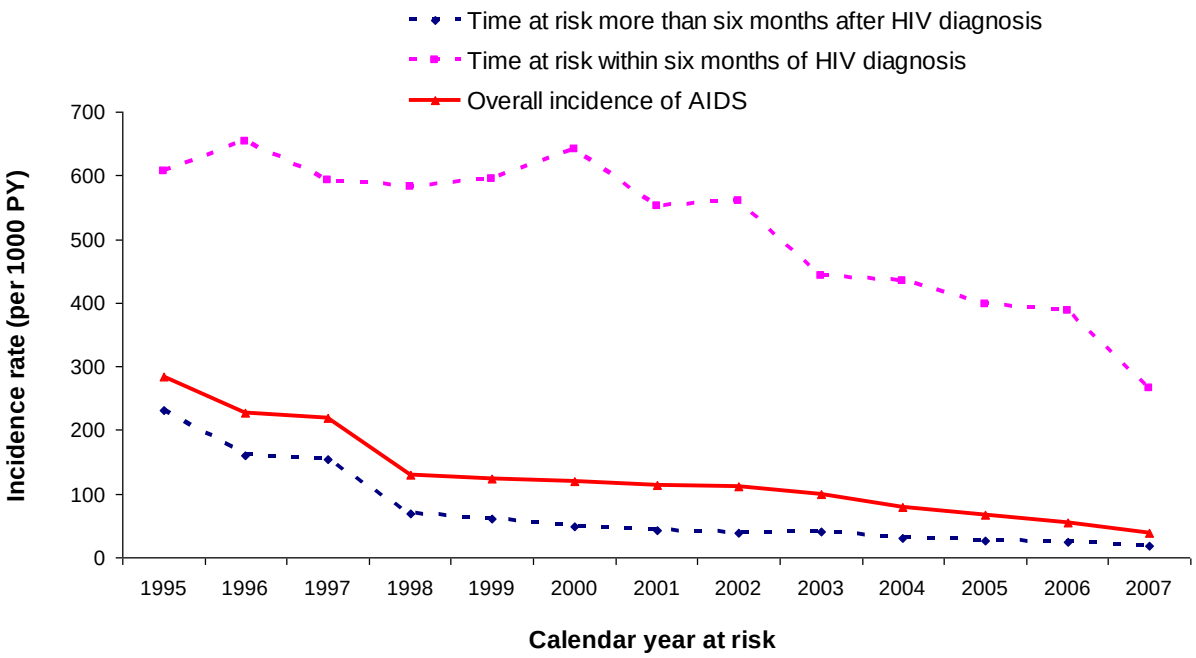
Figure 7.14. Incidence rates of AIDS by age



7.5.4 Trends in the incidence of AIDS

The incidence of AIDS decreased from 284.6 (267.1, 303.1) per 1,000 PYFU in 1995 to 38.4 (35.9, 41.0) per 1,000 PYFU in 2007 Table 7.4). There were marked declines between 1995 and 1998 and a gradual decline between 1998 and 2007. The decline in the incidence of AIDS within the first six months after HIV diagnosis was less marked than that seen for AIDS occurring more than six months after HIV diagnosis (Figure 7.15).

Figure 7.15. Trends in the incidence of AIDS stratified according to follow-up within, or more than, six months after HIV diagnosis



7.5.5 Poisson Regression

There were significant variations in the incidence rate ratios for AIDS in univariable analyses for each factor considered, similar to that observed for deaths (Table 7.5). The relationship between AIDS incidence and square root of the CD4 cell count was very similar to that for the incidence of death (IRR 0.79; 95% CI 0.78, 0.79). The incidence rate ratio for PYFU within the first six months after HIV diagnosis was greater than that for the incidence of death (IRR 10.45; 95% CI 10.09, 10.82). The incidence rate among individuals under follow-up in 2007 was about a eighth of that in 1995 (IRR 0.13; 95% CI 0.12, 0.15).

Adjustment for the CD4 cell count markedly decreased the decline in incidence observed between 1995 and 1998 and explained much of the change in the incidence of AIDS between 1995 and 2000 (bivariable analysis – Figure 7.16). However, the trend after 2000 was similar to that in the univariable model. Adjustment for factors other than the CD4 cell count made the trend over time slightly more marked in comparison to the univariable model. Adjustment for all factors in a multivariable analysis resulted in a decline that was more marked than the univariable analysis but less marked than the bivariable analysis. Multivariable analysis attenuated the incidence rate ratios for all factors studied except for age (IRR 1.16; 95% CI 1.14, 1.18 compared to IRR 1.05; 95% CI 1.03, 1.07 previously) (Table 7.5).

Figure 7.16. Incidence rate ratio of AIDS by calendar year at risk: comparison of different models

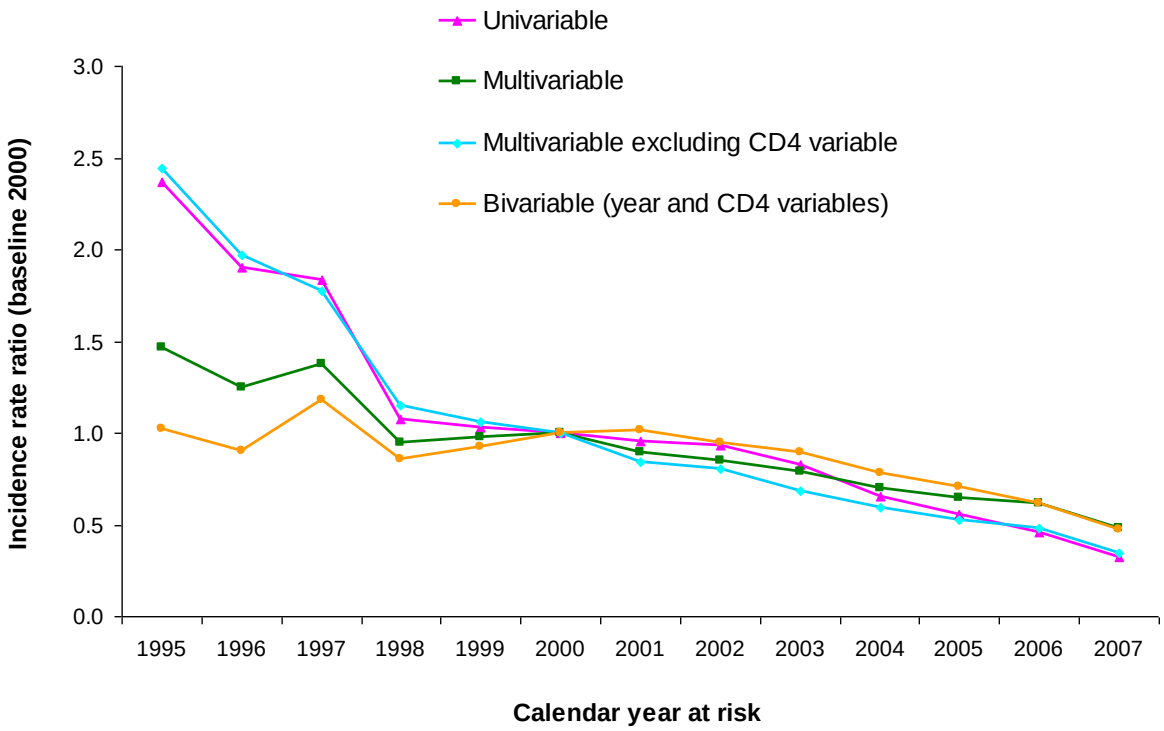


Table 7.5. Results from univariable and multivariable Poisson regression analysis of the association between each factor and the incidence of AIDS

		Univariable incidence rate ratio (95% confidence interval)	Multivariable incidence rate ratio (95% confidence interval)
Calendar year			
	1995	**2.37 (2.16, 2.60)**	**1.46 (1.33, 1.61)**
	1996	**1.90 (1.73, 2.09)**	**1.25 (1.13, 1.37)**
	1997	**1.83 (1.67, 2.01)**	**1.38 (1.26, 1.51)**
	1998	1.07 (0.97, 1.19)	0.95 (0.85, 1.05)
	1999	1.03 (0.93, 1.14)	0.98 (0.89, 1.08)
	2000	-	-
	2001	0.95 (0.87, 1.04)	**0.89 (0.81, 0.98)**
	2002	0.93 (0.85, 1.02)	**0.85 (0.78, 0.93)**
	2003	**0.83 (0.76, 0.90)**	**0.79 (0.72, 0.86)**
	2004	**0.66 (0.60, 0.72)**	**0.70 (0.64, 0.76)**
	2005	**0.56 (0.51, 0.61)**	**0.64 (0.59, 0.70)**
	2006	**0.46 (0.42, 0.50)**	**0.61 (0.56, 0.67)**
	2007	**0.32 (0.29, 0.35)**	**0.48 (0.44, 0.53)**
Square root of CD4 cell count		**0.79 (0.78, 0.79)**	**0.82 (0.82, 0.83)**

PYFU within first six months after HIV diagnosis		
No	-	-
Yes	**10.45 (10.09, 10.82)**	**5.79 (5.58, 6.02)**
Region		
Outside London	**1.12 (1.08, 1.16)**	**1.09 (1.05, 1.13)**
London	-	-
Age, per additional 10 years	**1.05 (1.03, 1.07)**	**1.16 (1.14, 1.18)**
Risk group		
MSM	-	-
Heterosexual men	**2.09 (1.99, 2.18)**	0.99 (0.93, 1.05)
Heterosexual women	**1.54 (1.48, 1.61)**	0.96 (0.91, 1.02)
Heterosexual women (diagnosed antenatally)	**0.48 (0.42, 0.55)**	**0.49 (0.42, 0.57)**
IDU	**1.63 (1.49, 1.80)**	0.99 (0.90, 1.09)
Recipients of blood products	1.15 (0.98, 1.35)	**0.61 (0.51, 0.71)**
Ethnicity		
White	-	-
Black African	**1.66 (1.60, 1.72)**	**1.11 (1.05, 1.18)**
Black Caribbean	**1.28 (1.15, 1.42)**	0.97 (0.87, 1.08)
Black Other	**1.20 (1.05, 1.38)**	0.89 (0.77, 1.02)
Indian\Pakistani\Bangladeshi	**1.90 (1.65, 2.17)**	**1.44 (1.25, 1.66)**
Other Asian	**1.76 (1.52, 2.02)**	**1.42 (1.23, 1.64)**
Other/mixed	1.04 (0.94, 1.15)	0.96 (0.87, 1.07)

7.5.6 Interaction between calendar year and follow-up within six months of HIV diagnosis

The analysis was repeated including an interaction between calendar year and whether follow-up was within six months of HIV diagnosis or not. The model indicated that the incidence of AIDS during follow-up within six months after HIV diagnosis increased between 1995 and 2000 (IRR 0.37; 95% CI 0.31, 0.45 in 1995 compared to 2000) but then remained largely stable to 2007 (IRR 0.97; 95% CI 0.80, 1.18 in 2007 compared to 2000) (Table 7.6). In contrast, the rate ratios for follow-up more than six months after HIV diagnosis followed a similar trend as in the overall multivariable analysis. Rate ratios compared to 2000

decreased markedly between 1995 (IRR 2.29; 95% CI 2.00, 2.63) and 1998 (IRR 1.17; 95% CI 1.00, 1.38) and then continued to decrease more gradually until 2007 (IRR 0.50; 95% CI 0.43, 0.58). Rate ratios for other factors did not appreciably change.

Table 7.6. Results from multivariable Poisson analysis of the association between each factor and the incidence of AIDS including an interaction between calendar year and PYFU within first six months after HIV diagnosis

	Multivariable incidence rate ratio (95% confidence interval)	**Multivariable incidence rate ratio: interaction between calendar year and PYFU within six months after HIV diagnosis (95% confidence interval)**
Square root of CD4 cell count	**0.82 (0.82, 0.83)**	**0.83 (0.83, 0.83)**
PYFU within first six months after HIV diagnosis		
No	-	-
Yes	**5.79 (5.58, 6.02)**	**7.53 (6.54, 8.67)**
Region		
Outside London	**1.09 (1.05, 1.13)**	**1.08 (1.04, 1.12)**
London	-	-
Age, per additional 10 years	**1.16 (1.14, 1.18)**	**1.17 (1.14, 1.19)**
Risk group		
MSM	-	-
Heterosexual men	**2.09 (1.99, 2.18)**	0.99 (0.93, 1.05)
Heterosexual women	**1.54 (1.48, 1.61)**	0.97 (0.91, 1.03)
Heterosexual women (diagnosed antenatally)	**0.48 (0.42, 0.55)**	**0.48 (0.42, 0.56)**
IDU	**1.63 (1.49, 1.80)**	1.01 (0.92, 1.11)
Recipients of blood products	1.15 (0.98, 1.35)	0.57 (0.49, 0.68)

		all PYFU	PYFU more than six months after HIV diagnosis
Calendar year			
	1995	**1.46 (1.33, 1.61)**	**2.29 (2.00, 2.63)**
	1996	**1.25 (1.13, 1.37)**	**1.71 (1.48, 1.98)**
	1997	**1.38 (1.26, 1.51)**	**2.12 (1.85, 2.44)**
	1998	0.95 (0.85, 1.05)	**1.17 (1.00, 1.38)**
	1999	0.98 (0.89, 1.08)	1.11 (0.95, 1.30)
	2000	-	-
	2001	**0.89 (0.81, 0.98)**	0.94 (0.80, 1.10)
	2002	**0.85 (0.78, 0.93)**	**0.83 (0.71, 0.97)**
	2003	**0.79 (0.72, 0.86)**	0.90 (0.78, 1.04)
	2004	**0.70 (0.64, 0.76)**	**0.71 (0.61, 0.83)**
	2005	**0.64 (0.59, 0.70)**	**0.66 (0.57, 0.77)**
	2006	**0.61 (0.56, 0.67)**	**0.61 (0.53, 0.70)**
	2007	**0.48 (0.44, 0.53)**	**0.50 (0.43, 0.58)**
			PYFU within six months after HIV diagnosis
	1995		**0.37 (0.31, 0.45)**
	1996		**0.55 (0.45, 0.67)**
	1997		**0.43 (0.36, 0.53)**
	1998		**0.70 (0.56, 0.86)**
	1999		0.81 (0.67, 1.00)
	2000		-
	2001		0.91 (0.75, 1.11)
	2002		1.03 (0.85, 1.24)
	2003		**0.81 (0.67, 0.97)**
	2004		0.97 (0.80, 1.16)
	2005		0.95 (0.79, 1.14)
	2006		1.04 (0.87, 1.26)
	2007		0.97 (0.80, 1.18)
Ethnicity			
	White	-	-
	Black African	**1.66 (1.60, 1.72)**	**1.10 (1.03, 1.17)**
	Black Caribbean	**1.28 (1.15, 1.42)**	0.97 (0.87, 1.08)
	Black Other	**1.20 (1.05, 1.38)**	0.88 (0.76, 1.01)
	Indian\Pakistani\Bangladeshi	**1.90 (1.65, 2.17)**	**1.44 (1.25, 1.65)**
	Other Asian	**1.76 (1.52, 2.02)**	**1.39 (1.21, 1.61)**
	Other/mixed	1.04 (0.94, 1.15)	0.96 (0.87, 1.06)

7.5.7 Sensitivity analyses

7.5.7.1 Consideration of only AIDS diagnoses reported to HARS

Estimates of incidence decreased substantially when the analysis was repeated with consideration of only AIDS diagnoses reported to HARS. Although the shape of the trend over time was similar (Figure 7.17) there was a much more marked decline in the multivariable incidence rate ratios (from 2.62; 95% CI 2.34, 2.93 in 1995 to 0.33; 95% CI 0.28, 0.38 in 2007 when compared to 2000) (Table 7.7).

In addition to change in trend over time, there was a much higher rate ratio for individuals followed within the first six months after HIV diagnosis than for those followed subsequently (IRR 11.31; 95% CI 10.68, 11.97) than in the original model (Table 7.7). There were also changes in the rate ratios for other factors: in this model IDU were significantly less likely than MSM (IRR 0.77; 95% CI 0.67, 0.88), and black Caribbeans were significantly less likely than white individuals (IRR 0.83; 95% CI 0.71, 0.98), to be diagnosed with AIDS. However, black Africans were no longer significantly more likely to be diagnosed with AIDS than white individuals (IRR 1.05; 95% CI 0.97, 1.14). Individuals receiving care outside London were even more likely to be diagnosed with AIDS than those cared for in London in comparison to the original model (IRR 1.31; 95% CI 1.24, 1.37).

Figure 7.17. Incidence rates and rate ratios of AIDS by calendar year of follow-up – results from sensitivity analyses.

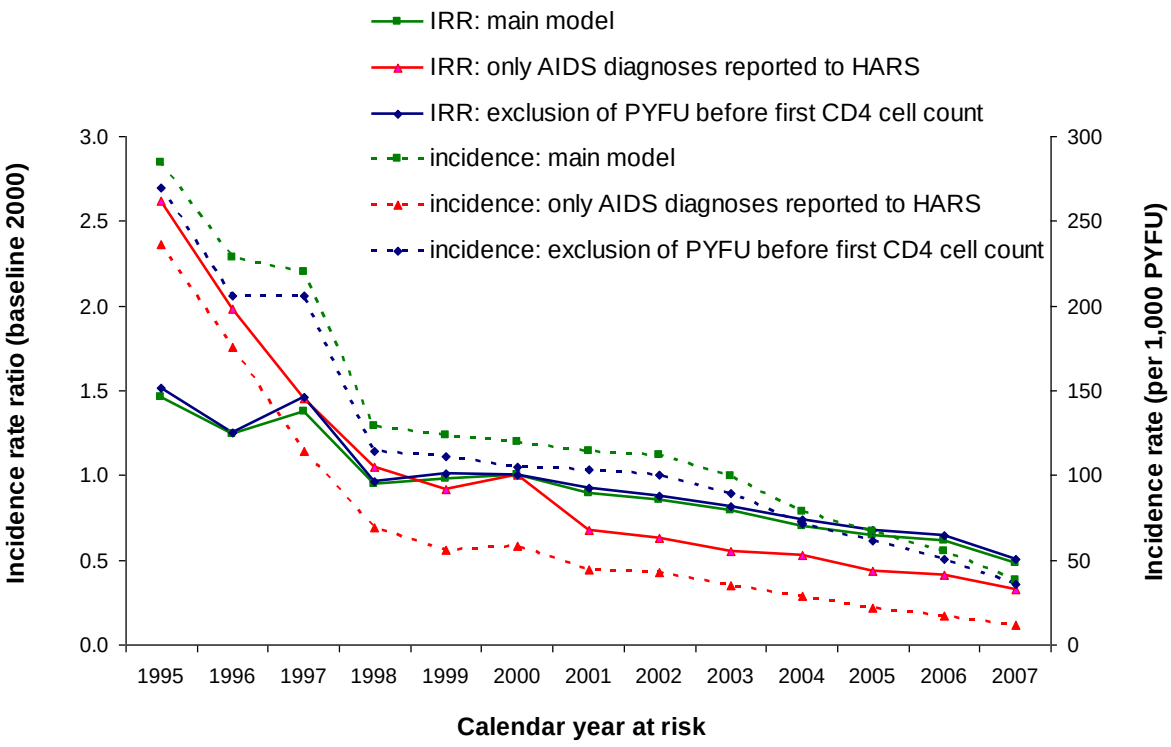


Table 7.7. Results of sensitivity analyses from multivariable Poisson regression analysis of the association between each factor and the incidence of AIDS

		Multivariable incidence rate ratio (95% confidence interval)		
		Original model	**Only considering AIDS from HARS**	**Excluding PYFU before first CD4 cell count**
Calendar year				
	1995	**1.46 (1.33, 1.61)**	**2.62 (2.34, 2.93)**	**1.51 (1.37, 1.67)**
	1996	**1.25 (1.13, 1.37)**	**1.98 (1.76, 2.23)**	**1.25 (1.13, 1.38)**
	1997	**1.38 (1.26, 1.51)**	**1.46 (1.29, 1.64)**	**1.46 (1.32, 1.61)**
	1998	0.95 (0.85, 1.05)	1.05 (0.92, 1.20)	0.96 (0.86, 1.07)
	1999	0.98 (0.89, 1.08)	0.92 (0.80, 1.05)	1.01 (0.91, 1.13)
	2000	-	-	-
	2001	**0.89 (0.81, 0.98)**	**0.67 (0.59, 0.77)**	0.92 (0.84, 1.02)
	2002	**0.85 (0.78, 0.93)**	**0.63 (0.56, 0.72)**	**0.88 (0.80, 0.97)**
	2003	**0.79 (0.72, 0.86)**	**0.55 (0.49, 0.63)**	**0.82 (0.75, 0.90)**
	2004	**0.70 (0.64, 0.76)**	**0.53 (0.47, 0.60)**	**0.74 (0.67, 0.81)**
	2005	**0.64 (0.59, 0.70)**	**0.44 (0.39, 0.50)**	**0.67 (0.61, 0.74)**
	2006	**0.61 (0.56, 0.67)**	**0.41 (0.36, 0.47)**	**0.64 (0.58, 0.71)**
	2007	**0.48 (0.44, 0.53)**	**0.33 (0.28, 0.38)**	**0.51 (0.46, 0.56)**

Square root of CD4 cell count	**0.82 (0.82, 0.83)**	**0.80 (0.79, 0.80)**	**0.82 (0.82, 0.83)**
PYFU within first six months after HIV diagnosis			
No	-		
Yes	**5.79 (5.58, 6.02)**	**11.31 (10.68, 11.97)**	**5.44 (5.22, 5.66)**
Region			
Outside London	**1.09 (1.05, 1.13)**	**1.31 (1.24, 1.37)**	**1.08 (1.04, 1.12)**
London	-		
Age, per additional 10 years	**1.16 (1.14, 1.18)**	**1.16 (1.13, 1.19)**	**1.15 (1.13, 1.18)**
Risk group			
MSM	-	-	-
Heterosexual men	0.99 (0.93, 1.05)	0.94 (0.86, 1.02)	0.96 (0.90, 1.03)
Heterosexual women	0.96 (0.91, 1.02)	0.95 (0.87, 1.03)	0.94 (0.88, 1.01)
Heterosexual women (diagnosed antenatally)	**0.49 (0.42, 0.57)**	**0.37 (0.28, 0.49)**	**0.52 (0.45, 0.61)**
IDU	0.99 (0.90, 1.09)	**0.77 (0.67, 0.88)**	0.99 (0.89, 1.09)
Recipients of blood products	**0.61 (0.51, 0.71)**	**0.59 (0.47, 0.75)**	**0.56 (0.47, 0.67)**
Ethnicity			
White	-	-	-
Black African	**1.11 (1.05, 1.18)**	1.05 (0.97, 1.14)	**1.12 (1.05, 1.19)**
Black Caribbean	0.97 (0.87, 1.08)	**0.83 (0.71, 0.98)**	0.94 (0.84, 1.05)
Black Other	0.89 (0.77, 1.02)	0.94 (0.77, 1.14)	0.89 (0.76, 1.03)
Indian\Pakistani\Bangladeshi	**1.44 (1.25, 1.66)**	**1.41 (1.16, 1.70)**	**1.49 (1.28, 1.72)**
Other Asian	**1.42 (1.23, 1.64)**	**1.62 (1.34, 1.96)**	**1.36 (1.16, 1.59)**
Other/mixed	0.96 (0.87, 1.07)	1.06 (0.92, 1.23)	0.97 (0.87, 1.08)

7.5.7.2 Exclusion of PYFU between HIV diagnosis and first CD4 cell count

Estimates of incidence were slightly lower throughout the period when the analysis was repeated with exclusion of PYFU between HIV diagnosis and first CD4 cell count (Figure 7.17). However, the incidence rate ratios for the various factors were inconsequentially different than those in the original model (Table 7.7).

7.5.7.3 Termination of all follow-up at the end of 1999

Estimates of incidence were very similar in 1996, 1997 and 1998 to those in the original model when the analysis was repeated with termination of all follow-up at the end of 1999. However, incidence was estimated to be higher in 1995 and 1999 in this model. The multivariable analysis reflected this as the estimated rate in 1998 was lower than that in 1999 (IRR 0.81; 95% CI 0.73, 0.89) and the rate for 1996 was no longer significantly higher than that in 1999 (IRR 1.07; 95% CI 0.98, 1.18). The rate for heterosexual women diagnosed antenatally was no longer significantly different to that for MSM, likely due to small numbers before 2000 due to under-ascertainment of pregnancy status (IRR 0.71; 95% CI 0.41, 1.24). The latter may also have been the cause for other heterosexual women to be significantly less likely than MSM to be diagnosed with AIDS in this model (IRR 0.83; 95% CI 0.74, 0.93). Differences in the rate ratios between the models for the other factors were minor.

7.6 Discussion

The overall incidence of death was 18.5 per 1,000 PYFU and the overall incidence of AIDS was 93.0 per 1,000 PYFU in E,W&NI between 1995 and 2007. There were marked declines in the incidence of both AIDS and death through the early HAART era, which then continued more slowly between 1998 and 2007. These declines in incidence reflected the population effect of combination therapy with nucleoside reverse transcriptase inhibitors and protease inhibitors[365;449]. After taking improvements in CD4 cell counts and other factors into account, reductions in mortality and AIDS incidence were less marked between 1995 and 1998, though still statistically significant. The incidence of AIDS and death changed little between 1998 and 2000 but

declined again after 2000/2001. The reasons for these trends cannot be determined from this population-based observational study. Suggested hypotheses may be the introduction of boosted protease inhibitors and new classes of ART, by providing new effective treatment options for patients, the availability of simpler ART regimens, by supporting improved adherence, or increasing potency with time since initiation[450]. Also, changes in reporting in 2000, which improved HIV reporting detail, may have reduced AIDS reporting.

Importantly, the incidence of both AIDS and death was continually high at low CD4 cell counts and the absolute PYFU with CD4 cell counts below 100 cells/mm^3 changed little over the period. This was despite marked declines in the proportion of overall PYFU spent with CD4 cell counts below 100 cells/mm^3 (falling from 26.0% in 1995 to 2.7% in 2007 in the analysis of death, and from 12.9% in 1995 to 1.7% in 2007 in the analysis of AIDS). An increasing proportion of these PYFU were contributed by individuals recently diagnosed and might be reduced by earlier HIV testing (Chapter 6).

These results show the high burden of AIDS and death at the time of HIV diagnosis. AIDS within the first six months of HIV diagnosis accounted for half of the total number of AIDS and deaths within the first six months of HIV diagnosis accounted for one in seven of the total number of deaths. After adjusting for other factors, the overall incidence of AIDS within six months of diagnosis was almost six times higher than that among patients followed more than six months after diagnosis. The equivalent figure for the incidence of death was 35%. However, the respective incidence rates were seven and a half times and 54% higher overall when taking into account the interaction between

calendar year and PYFU within the first six months after HIV diagnosis. These results fit with those of the analyses of late diagnosis and consequent mortality (Chapter 6). Other cohort studies have acknowledged this association because they have excluded patients with low CD4 cell counts at enrolment or the period shortly after diagnosis to remove their effect on the results of analyses[451-453]. However, there do not appear to be any published studies that have focused on the period shortly after HIV diagnosis as a particular period of high risk of AIDS or death for comparison. The explanation for a higher risk of AIDS or death within the first six months after HIV diagnosis cannot be known from these data. It can be hypothesised that this may be due the impact of treatment and care, which do not always start immediately after diagnosis and were not adjusted for in this population-level analysis. This would include prophylactic treatment such as the antibacterial cotrimoxazole, antiretroviral treatment and its impact on viral load, and routine clinical review to detect and intervene against opportunistic infections and other clinical conditions.

There was also little evidence of published literature comparing the population incidence of AIDS and death among HIV-infected individuals between ethnic groups after controlling for risk groups[454]. Studies have shown similar survival rates for black/African and white/European individuals but have not simultaneously considered the effect of risk groups[79-82;455]. These results indicate that ethnicity seems to have had little effect on mortality after adjusting for other determining factors. However, black Africans, Indian/Pakistani/Bangladeshi individuals and other Asians experienced higher AIDS incidence than white individuals in E,W&NI between 1995 and 2007 even after adjusting for other factors. This may have been due to higher background

rates of AIDS defining illnesses and particularly tuberculosis among these population groups[207;456]. Further investigation of the incidence of specific AIDS-defining illnesses by ethnic group is warranted but may be limited in E,W&NI by the under-reporting of AIDS to HARS. This appeared to be a particular limitation for AIDS diagnoses that occurred after the time of HIV diagnosis and is likely to be a consequence of the integrated CHR form and a failure of some clinicians to appreciate that AIDS is still a useful epidemiological marker, even if less important clinically. Overall estimates of the incidence of all first AIDS diagnoses were available through integration of data from SOPHID but specific AIDS-defining illnesses were not available from this surveillance system. Restriction of these analyses to only first AIDS diagnoses minimised the need to consider competing risks, but death remains a competing risk for first AIDS[457].

Mortality was substantially different between risk groups. There was lower mortality among heterosexuals compared to MSM even after controlling for all other factors. This may suggest that not all of the pertinent determining factors were included in this analysis, such as ongoing access to care and differences in uptake of ART between groups. However, other studies have not had data on ethnicity or sufficiently large ethnic groups to detect differences between risk groups after adjustment for ethnicity. The low mortality of pregnant women is likely to be due to them having been diagnosed early due to antenatal screening. This effect has not previously been shown to be an added benefit to the reduced transmission of infection from mother to infant and may be further strong evidence for the expansion of HIV testing among relatively low-risk populations (Section 6.5). As other studies have shown, IDUs had a higher mortality than all other groups (more than double that of MSM)[458-460]. Other

studies have shown that this is largely due to non-HIV related causes of deaths among IDUs[461-464].

AIDS incidence, in contrast to mortality, was only lower for heterosexual women diagnosed antenatally and recipients of blood/blood products than for MSM. The reasons for this cannot be determined from this analysis but the routine monitoring of the health of these individuals may potentially have played some part in the prevention of illness among these groups.

The results indicate similar trends in the incidence of AIDS and death [365;451;457;465-473] and a similar effect of CD4 cell counts on the incidence of AIDS and death[360;365;451;471;473-479] as large observational cohorts. The results also describe the effects of demographic factors. Incidence of both AIDS and death fell by about 20% for each doubling in CD4 cell counts. Each 10 year increase in age was associated with a 37% increase in mortality and a 16% increase in AIDS incidence. Also, previous AIDS diagnoses were associated with more than a doubling in mortality rates as previously reported[450;480].

CD4 cell counts are very strong markers for the risk of AIDS and death but improvements in CD4 cell counts over time did not account for the entire declines in incidence as noted in other studies. One hypothesis put forward is that HAART has a clinical benefit over and above that measured by the usual laboratory markers[365;474;478;480]. Others have stated that residual confounding due to measurement error in laboratory markers may explain this observation [481]. Allocation of PYFU to CD4 counts can also lead to variations in incidence rates though these have been shown to be minor[448]. Another limitation of this analysis is that laboratory markers other than CD4 cell counts were not

considered as they were not available from surveillance data. Although longitudinal viral loads could be considered in the future now that they are reported to SOPHID, this may not affect the results significantly as other studies have concluded that the inclusion of viral loads and other laboratory markers in models of incidence was unlikely to substantially affected trends[365].

This analysis shows that surveillance data can be used to create a pseudo-cohort with very large numbers of patients (nearly 50,000) and PYFU (over 200,000 years) comparable to analyses of pooled data from multi-centre cohorts[250;449;452;470;471;476;482-491]. The use of CD4 cell counts to define person-years ensured that patients were under follow-up, which may otherwise be a limitation of using surveillance data. The cut-off of 365 days accounted for the small proportion of individuals who were temporarily LTFU for longer (Section 5.4.3.3) and ensured that follow-up time was appropriately allocated to a level of risk/CD4 cell count. Changing the censoring definition of temporary LTFU had little effect on the trends in mortality observed. Indeed, very similar results were obtained from all sensitivity analyses on the mortality data, indicating that interpretations were robust. There was negligible effect of the use of CD4 cell count categories or either the square root or logarithm of CD4 cell counts on the trends over time. And, although the total number of PYFU and deaths was reduced, when either only the first continuous phase of follow-up was considered or when follow-up after 59 years of age was excluded, this had little effect on the overall estimates of incidence or the rate ratios. Analysis of mortality with follow-up split one year after HIV diagnosis instead of after six months, reduced the rate ratio for PYFU during this period in comparison to

subsequent follow-up because the incidence of death fell rapidly after HIV diagnosis due to the impact of treatment.

There was a marked decline in the proportion of first AIDS diagnoses that were reported to HARS. HARS reports accounted for a large majority of records with AIDS diagnoses in 1995 but only a third of all such records in 2007 and the trend was greater for AIDS diagnoses that occurred more than six months after an HIV diagnosis. The change in reporting methods in 2000 and the lower emphasis of the importance of reporting AIDS diagnoses[492] appear to have reduced the ascertainment of the HARS surveillance system. Estimates of incidence that excluded SOPHID reports of AIDS were less than half that when all reports were used (40.1 compared to 93.0 per 1,000 PYFU). Additionally, AIDS was only collected in SOPHID reports since 2005, and therefore much of these data were retrospectively collated (although prospectively collected locally in patient notes) and would be incomplete for individuals not reported to SOPHID since 2005 due to either death or LTFU. Although the shape of the trend over time was similar between models, adjustment for CD4 cell counts and other factors had little effect on this shape when only HARS reports of AIDS were considered, which may be due to under-reporting of AIDS to HARS. There were also some significant differences in the rate ratios for other factors when only AIDS diagnoses reported to HARS were considered. In particular, the effect of PYFU within the first six months of HIV diagnosis was exaggerated, but also IDU, black Caribbeans and individuals accessing care in London were relatively less likely to be reported as having AIDS.

I attempted to link all CD4 cell counts for the same patient even if these were reported from different laboratories but some lack of identifiers may have resulted in gaps in the dataset. The linkage of records from different people and the inaccurate recording of information should have been minimised in the development of the integrated dataset and was not believed to have resulted in any significant bias because follow-up time, AIDS and deaths were aggregated at a specific level of risk/CD4 cell count. Inclusion bias may have substantially affected the results because individuals included in the integrated dataset were less likely to have AIDS or die than other individuals, and particularly at the time of HIV diagnosis. In particular, the lower incidence of death among heterosexuals may be an artefact, there may have been an undetected higher incidence of AIDS among heterosexuals than among MSM, and the higher incidence of AIDS among black Africans may have been underestimated. Censoring may have affected the results for 1995 and 2007 particularly. For example, there was some evidence of an increase in mortality from 1995 to 1996, which may have been due to biased data for 1995. This may also have exaggerated the decline between 2006 and 2007.

Surveillance information about HAART, HIV-related but non-HAART treatment and prophylaxis was limited at the time of these analyses, which therefore could not directly account for the independent effect of HAART. However, this population should have similar access to therapy (as recommended by national guidelines[133]) and therefore comparisons of incidence rates are justified in this population-level analysis adjusted for period of HIV diagnosis[493].

7.7 Conclusion

In conclusion, the incidences of AIDS and death declined substantially in E,W&NI from the pre-HAART to the HAART era and continued to decline throughout the HAART era. However, individuals diagnosed late benefited less and increasingly experienced disproportionate morbidity and mortality. The integrated dataset demonstrated that changes in the surveillance of AIDS in E,W&NI since 2000 decreased the completeness of AIDS diagnoses in HARS (especially those occurring after HIV diagnosis) and this limited analysis of the incidence of specific AIDS-defining illnesses. However, analysing the overall incidence of AIDS and the incidence of deaths using integrated surveillance data can be an important public health tool for monitoring the population effectiveness of treatment and complements randomised clinical trials of the individuals effectiveness of treatment[493].

Chapter 8. Estimating ART start dates

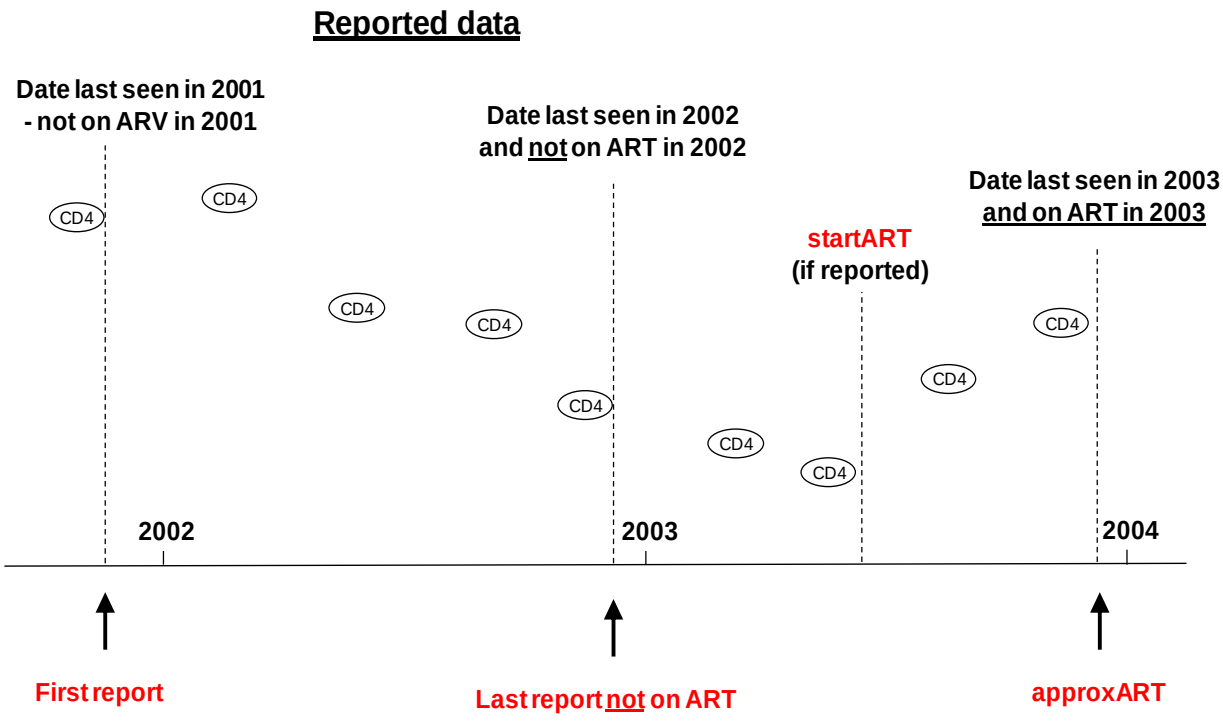
8.1 Introduction

Determining the date that an HIV-infected individual started ART (startART) is vital when analysing longitudinal data, such as the integrated dataset, because the progression of infection changes dramatically. Generally, viral loads decrease, CD4 cell counts increase, and the risk of progression to AIDS and death decrease with sufficient adherence and in the absence of viral resistance. The date of starting ART is essential for determining whether patient follow-up occurred whilst the patient was on or off therapy but is not available in the HARS or CD4 Surveillance datasets. For example, CD4 cell counts used to calculate rates of decline before ART and rates of increase after ART need to be separated accordingly to minimise misclassification and underestimation of trends[485;494]. Good knowledge of the date of starting ART can also be used to identify the CD4 cell count just prior to starting treatment and therefore investigate whether individuals start treatment as recommended and/or whether rates of CD4 increase are determined by these baseline CD4 counts[485;495-500].

StartART has been requested in the SOPHID survey since 2005. However, these data were incomplete and some sites had difficulty in extracting information from their records. Also, collection of startART could not include individuals who had died, left the country or stopped attending HIV services before 2005. This is problematic because these individuals may have been more likely to start ART at low CD4 counts, which could bias results.

Information about when people start ART is also available as a combination of the date last seen for HIV care during a SOPHID survey period and the level of ART (number of independently-acting drugs) at that time. The earliest of these dates with ART reported, provides the earliest date in a calendar year when a patient can be assumed to be on ART (approxART) (Figure 8.1). It is probable that a patient started ART between the start of the survey period and approxART because it is unlikely that an individual started ART during a previous survey period without that being reported. However, approxART could not be determined for individuals on therapy in 1997 or 1998 with reports from previous years because the level of ART was not collected in 1995 or 1996 and appeared to be incomplete in 1997, the first year of collection of these data (for 59% of those with approxART in 1998 and startART reported, startART occurred before 1998 indicating that approxART had not been reported in 1997 [Section 8.4.1]).

Figure 8.1. Schematic showing the information reported to SOPHID



In this chapter, CD4 cell counts in the integrated dataset were investigated to determine whether they could be used to determine a reasonably accurate estimate of the date of starting ART (estART) that was closer to startART than approxART (therefore occurring between the start of the survey period and approxART). CD4 cell count measurements were used because i) they were the only source of information available from surveillance data that were more regular than SOPHID surveys, and ii) there is an overall downwards trend in CD4 cell counts among untreated patients, which is generally reversed following the initiation of treatment[485]. It was recognised that although immunological response to ART is usually rapid, CD4 counts may take some time to increase or may even decline after the start of effective treatment[364;499;501-503]. There is also substantial variability in CD4 counts such that slopes can fluctuate over the short-term while showing a smoother long-term trend.

The goal of the work described in this chapter was to evaluate the feasibility of developing an algorithm that could be applied to the integrated dataset to estimate estART for individuals without a reported startART. Potential algorithms were considered that defined estART based on an initial increase in CD4 counts and/or a change in slope from negative to positive.

8.2 Aims

a) To investigate a number of possible algorithms, based on changes in CD4 cell counts, to enable a simple, practical, and reasonably accurate estimation of estART for individuals known to have started ART in any particular year (known approxART), but for whom the exact date (startART)

was unknown. The date of each CD4 cell count was considered to potentially be a close estimate of the date of startART.

b) To validate the most promising algorithms using reported startART and to use this information to select the optimal algorithm.

c) To use the optimal algorithm to determine estART for individuals without a reported startART and use it to supplement the integrated dataset.

8.3 Methods

The methods evaluated for determining estART were a set of algorithms developed to detect changes in CD4 cell count measurements that often occur first at the time of starting ART. These were: a nadir; CD4 cell counts below the level at which guidelines recommend that treatment should be started; an increase in CD4 cell counts; a preceding decrease and subsequent increase in CD4 cell counts; a positive slope in CD4 cell counts; and a preceding negative slope and subsequent positive slope in CD4 counts. A simple smoothing technique was used to assess whether the removal of some natural or measurement variability would improve the accuracy of the algorithm.

EstART was defined as the date on which the CD4 cell count, selected by the algorithm, was measured. Only dates of CD4 count measurements earlier than approxART but in the same year were considered for the determination of estART because it was unlikely for patients to have started ART during a previous survey (which must not have been reported).

8.3.1 Data preparation

Where more than one startART was reported for an individual (383) the earliest reported startART was used for analysis (assuming individuals had changed clinics and each clinic had only reported the date they had started ART). For individuals on treatment but without startART reported, the date last seen for care (approxART) when first reported to be prescribed ART was considered to be the earliest date that an individual was known to be on ART (Figure 8.1).

Initial investigation of the integrated dataset showed that some individuals had a reported startART date after approxART. This was assumed to be either due to coding errors or a change of clinics with earlier clinics not reporting startART. Therefore, dates of startART after approxART were ignored in this analysis.

8.3.2 Defining and evaluating the algorithms

For each individual, the date of the first CD4 count to meet specified criteria for each algorithm was used to define estART. For four of the algorithms these criteria required a change in CD4 counts above a threshold, which was varied to determine the optimal level. An accurate estART was defined as one within 31 days of startART (the median time between startART and the preceeding CD4 count was 21 days (Section 8.4.2)). A random sample of 70% of individuals was used to develop the algorithms, with the remainder used for validation[504].

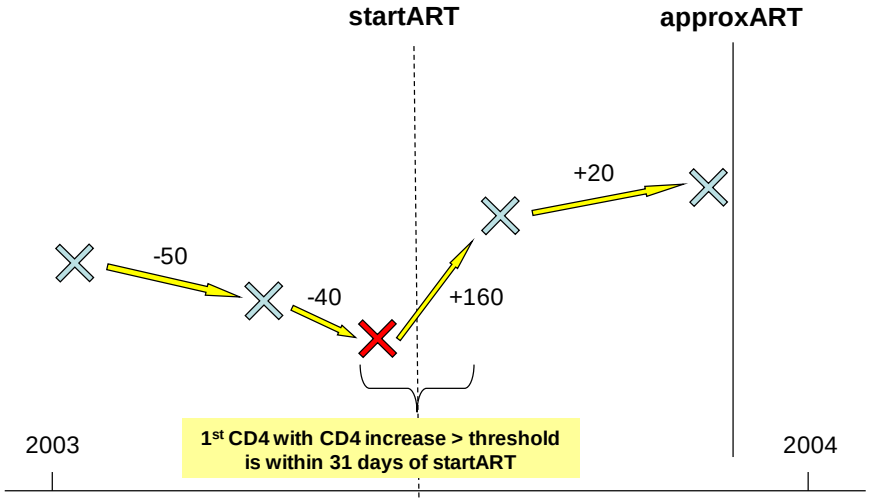
First, the algorithms were described by considering CD4 cell count changes at startART. This provided descriptive information to develop criteria for each algorithm. Then, cut-off values were tested for their sensitivity and specificity in categorising estART. For each cut-off value:

- true positives had estART within 31 days of startART (Figure 8.2a)
- false positives were where estART was not within 31 days of startART (Figure 8.2b)
- true negatives were where no estART could be estimated and no CD4 counts measured within 31 days of startART (Figure 8.2c)
- false negatives were where no estART could be estimated but there were CD4 counts measured within 31 days of startART (Figure 8.2d).

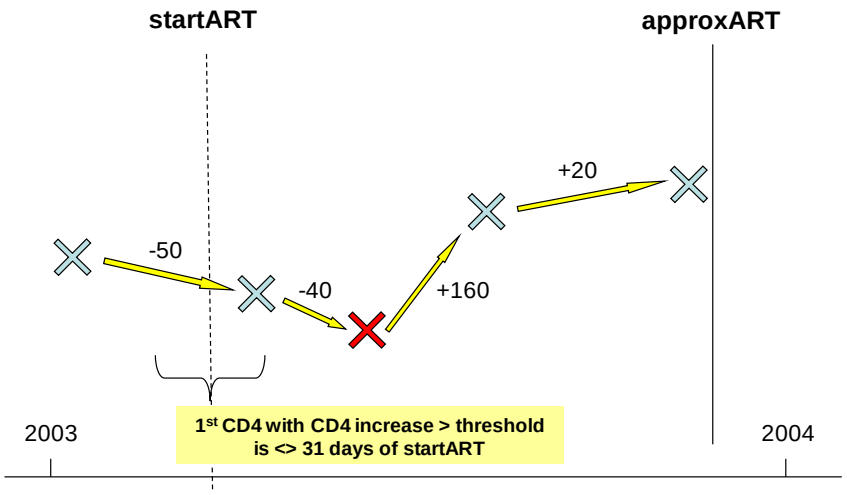
Summary tables of sensitivity, specificity and likelihood ratios were created to assess the performance of different cut-off values for each algorithm. The cut-off with the highest likelihood ratio was considered as the threshold likely to provide the best combination of sensitivity and specificity.

Figure 8.2. a, b, c, d). Examples of how patients were determined to be true positives, false positives, true negatives and false negatives:

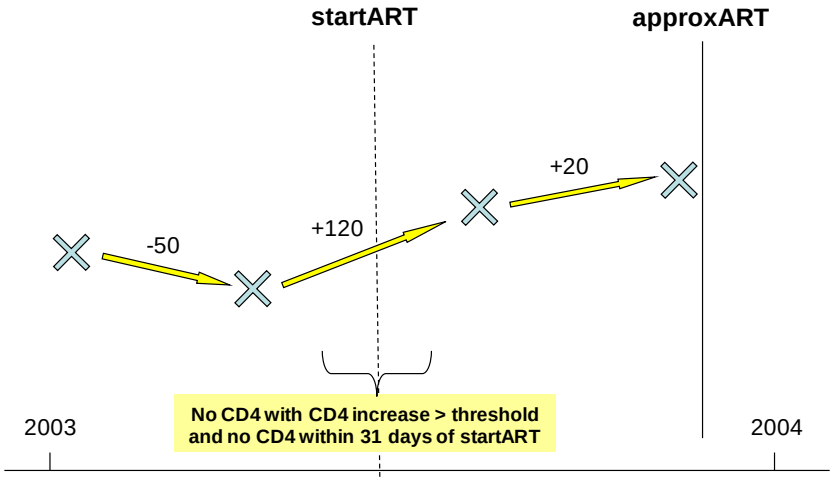
a) A true positive was an individual whose first CD4 count with a subsequent increase greater than the threshold was measured within 31 days of startART. In this example, the red cross depicts the first CD4 count with a subsequent increase (the yellow arrow) of greater than 150 cells/mm^3. That CD4 count is within 31 days date of startART (the bracket) so the patient was considered to be a true positive.



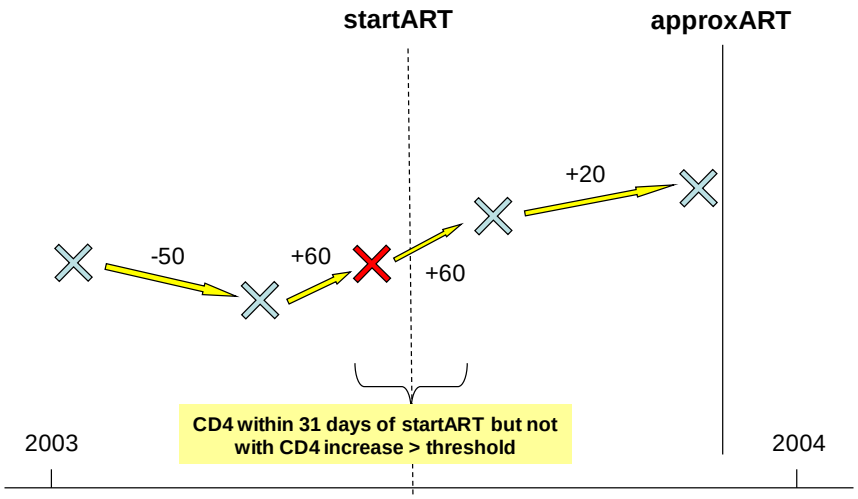
b) A false positive was an individual whose first CD4 count with a subsequent increase greater than the threshold was not measured within 31 days of startART. In this example, the patient does have an increase greater than the threshold (red cross and subsequent yellow arrow) but this does not fall within 31 days of startART (bracket) so the positive result falsely identifies startART.



c) A true negative was an individual who had no CD4 counts with a subsequent increase greater than the threshold and no CD4 counts measured within 31 days of startART. In this example without the CD4 count shown by the red cross, there is no CD4 count with a subsequent increase greater than 150 cells/mm^3 and no CD4 counts measured within 31 days of startART. Therefore, it is true to say that there are no CD4 counts that correctly identify startART.



d) A false negative was an individual with a CD4 count measured within 31 days of startART but without any CD4 counts with a subsequent increase greater than the threshold. In this example, the red cross depicts a CD4 count within 31 days of startART, which would be expected to be the first CD4 count with a subsequent increase greater than the threshold of 150 cells/mm^3. However, the subsequent increase of 60 cells/mm^3 is less than the threshold and therefore, falsely, no startART is identified for this patient.



8.3.3 Algorithms

Seven algorithms were considered, which could simply and practicably be applied to a large proportion of individuals in the integrated dataset. As algorithms were based on changes in CD4 counts, they could only be applied to individuals for whom a sufficient number of CD4 cell counts had been reported. This number varied between the algorithms and is described below (Section 8.4.3 and Table 8.2). For algorithms 3-7,

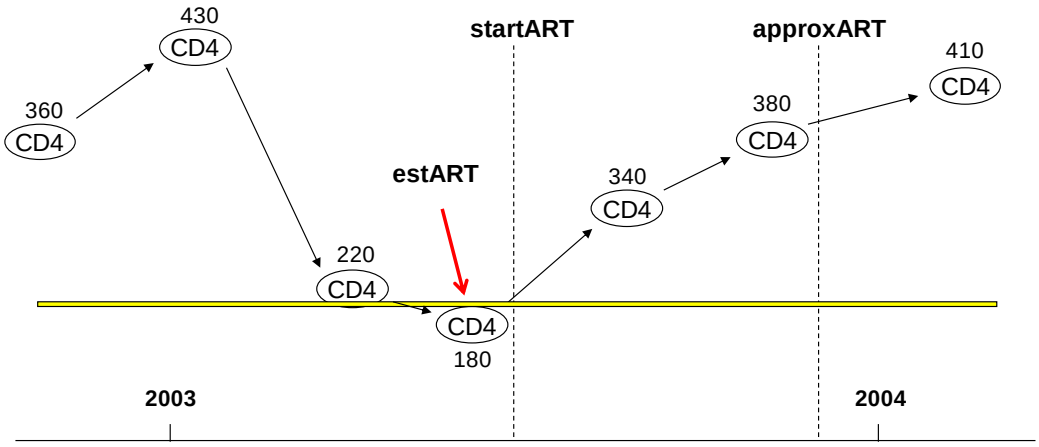
The algorithms evaluated were as follows:

1. The first time an individual meets the recommended criteria for starting ART: CD4 threshold or AIDS

   Published guidelines recommending the initiation of ART based on CD4 cell counts among people with established HIV infection have changed over time (Section 1.3.6). Guidelines have also recommended that people with primary HIV infection should be included in clinical trials, that people with symptomatic HIV infection or AIDS should start ART and that other factors in addition to CD4 cell count should be taken into account (including viral load, rate of CD4 cell count decline and age)[123].

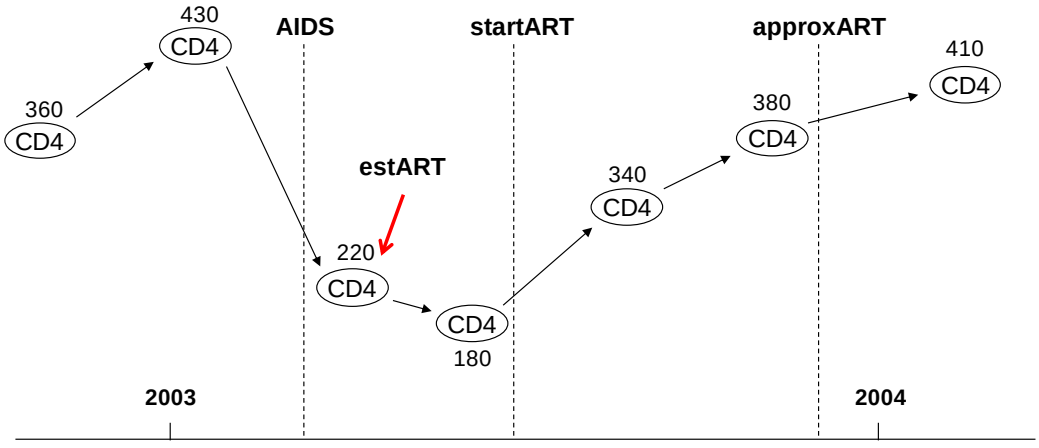
   In light of these guidelines, a CD4 cell count below a threshold (350 cells/mm^3 in 1999 and 2000 and 200 cells/mm^3 between 2001 and 2007) or a recent AIDS diagnosis was used to identify the first CD4 cell count in the year of approxART that should have prompted ART initiation[130-135;505].

a. EstART was determined as the first CD4 cell count below the recommended threshold for starting ART

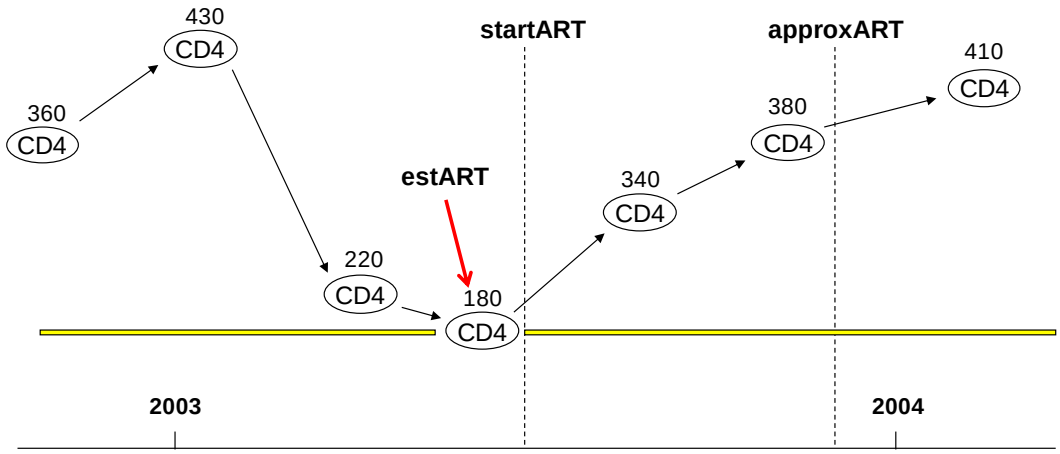


or

b. EstART was determined as the first CD4 cell count after a reported AIDS diagnosis (within the same or preceding month)

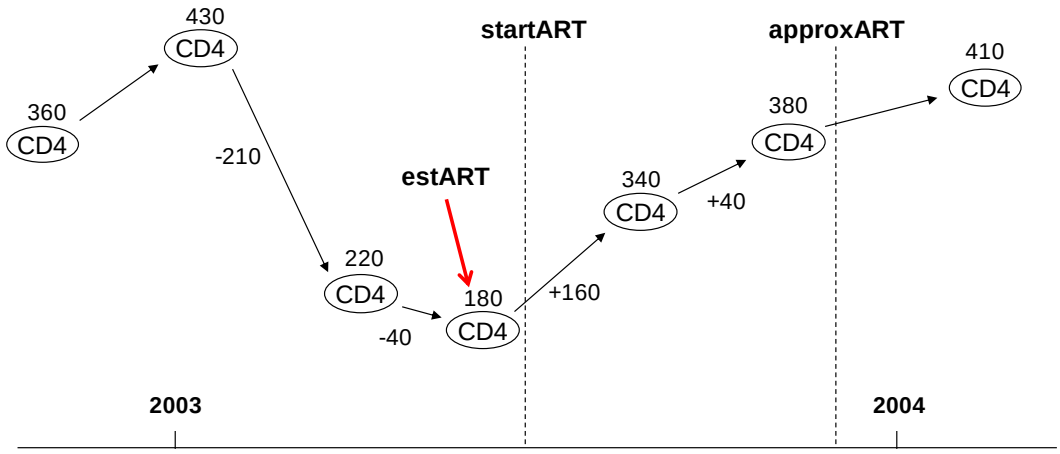


2. The date of the CD4 nadir (lowest CD4 cell count)



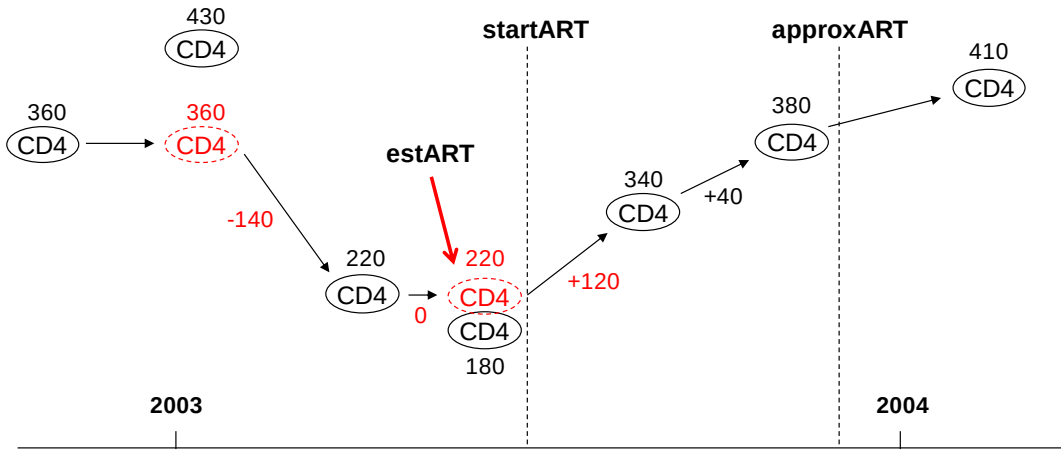
EstART was determined as the date of the lowest CD4 cell count, as long as it was earlier than, but in the same year as approxART (the earliest date was selected if there were two or more measurements with the same values).

3. An absolute increase immediately following a CD4 count



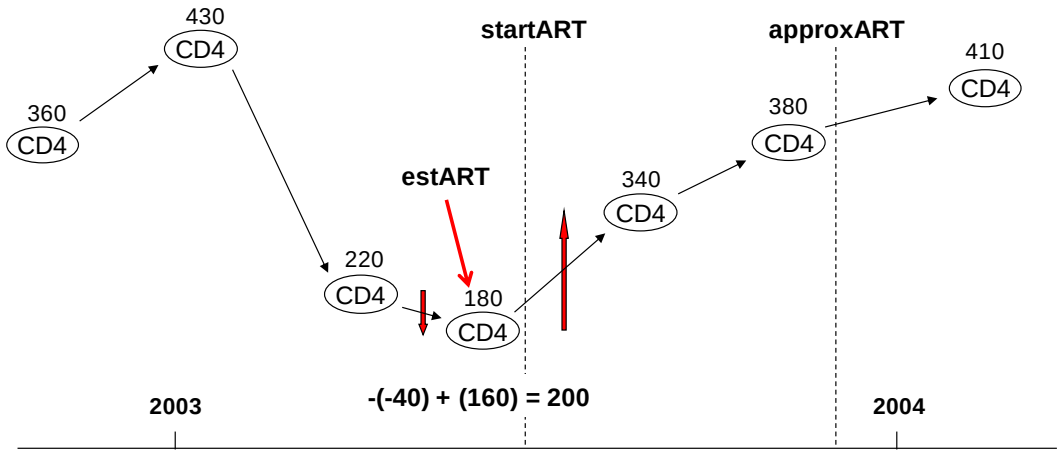
EstART was determined as the earliest CD4 cell count measurement where the difference between two consecutive CD4 cell counts exceeded the threshold.

4. An absolute increase immediately following a CD4 count after using a smoothing mechanism to remove some variability



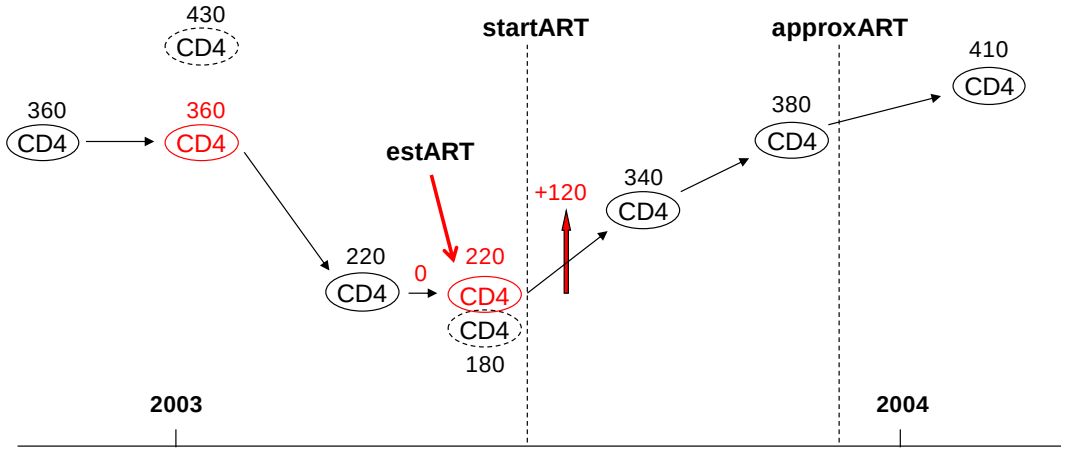
The running median of each three consecutive CD4 cell counts was used instead of the reported CD4 count to remove some variation. In this example, the CD4 count of 430 cells/mm^3 (dotted) and subsequent fall of 210 cells/mm^3 was replaced by a smoothed CD4 count of 360 cells/mm^3 (in red) and a drop of 140 cells/mm^3 (highlighted). Similarly, the CD4 cell count of 180 cells/mm^3 (dotted) was replaced by a smoothed count of 220 cells/mm^3 (in red), indicating no change before it and an increase of 120 cells/mm^3 after it (instead of -40 cells/mm^3 before and +160 cells/mm^3 after). estART was determined as the earliest date where the difference between two consecutive smoothed CD4 cell counts exceeded the threshold.

5. A preceding decrease and subsequent increase in CD4 counts



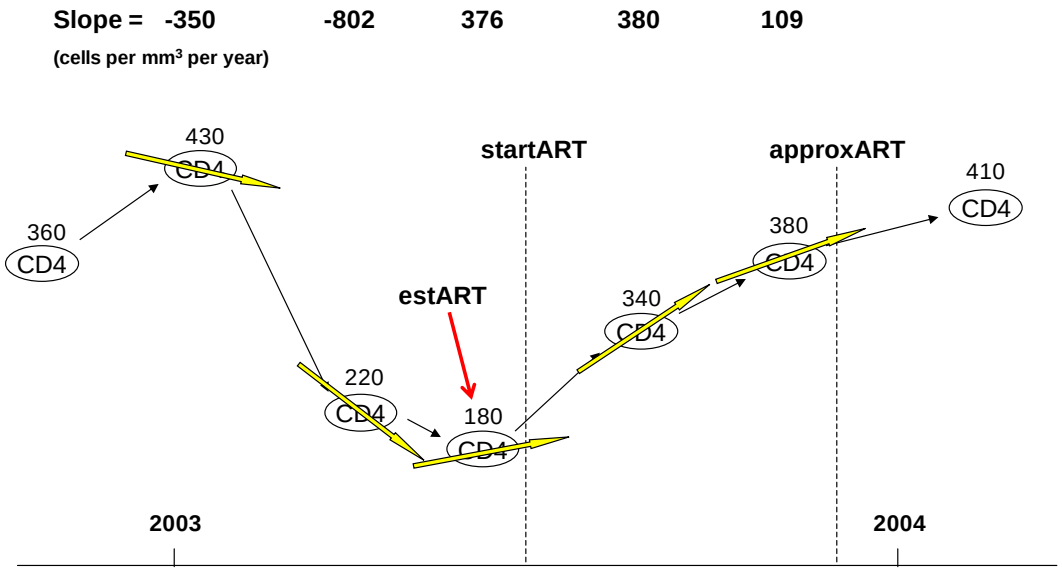
EstART was defined as the earliest CD4 count that was immediately preceded by a decrease and followed by an increase, such that the combined magnitude of change around the central measurement exceeded the threshold.

6. A preceding decrease and subsequent increase in smoothed CD4 counts



CD4 counts were smoothed using the median of three consecutive counts as before (dotted counts replaced). Where this occurred, the change between the smoothed count and one of the adjacent CD4 counts will be zero. EstART was determined as the earliest CD4 count that was preceded by no change and followed by an increase that exceeded the threshold.

7. A positive slope in CD4 counts



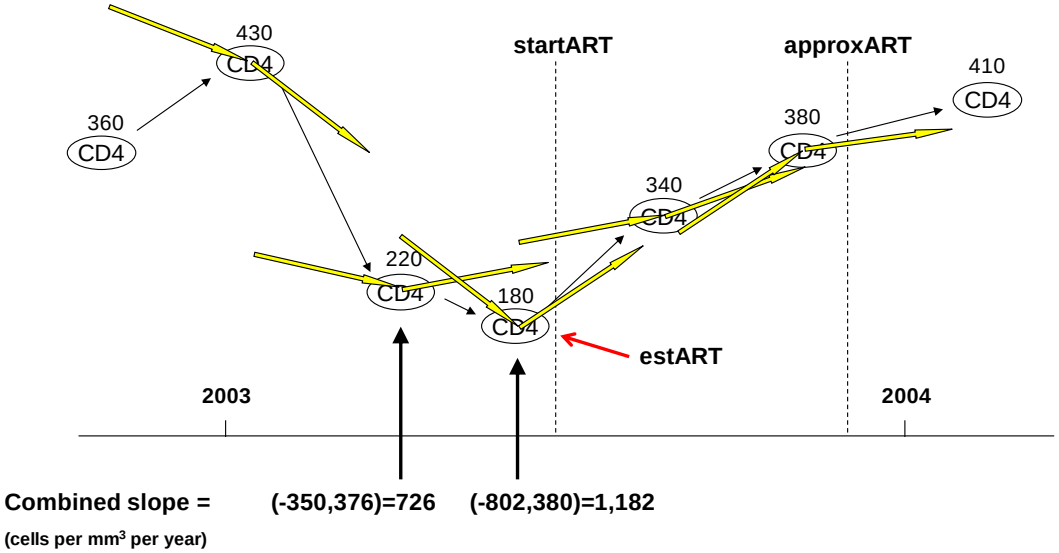
Both the adjacent CD4 cell counts were considered to determine the slope through each set of three consecutive CD4 cell counts (using the least squares method – see equation below). EstART was determined as the earliest CD4 count with a positive slope that exceeded the threshold.

$$slope = \frac{\sum_{i=1}^{3}(CD4_i - \overline{CD4}) \times (CD4date_i - \overline{CD4date}) / 365.25}{\sum_{i=1}^{3}((CD4date_i - \overline{CD4date}) / 365.25)^2}$$

Key:

Consecutive CD4 cell counts	$CD4_i$
Dates of CD4 cell count measurements	$CD4date_i$
Mean of consecutive CD4 cell counts	$\overline{CD4}$
Mean of CD4 cell count dates	$\overline{CD4date}$

8. A preceding negative slope and subsequent positive slope in CD4 counts



The figure shows the slopes through each set of three consecutive CD4 cell counts (around the specific CD4 count itself). EstART was determined as the earliest CD4 count with a preceding negative slope and subsequent positive slope with the sum of the magnitude of those slopes exceeding the threshold.

8.4 Results

8.4.1 Describing startART and approxART

Of 50,167 patients in the integrated dataset, there were 28,095 (56.0%) individuals who had startART reported and 34,974 (69.7%) with approxART (Figure 8.3). Of the 25,263 individuals with startART equal to or earlier than approxART:

- startART was in the same year as approxART for 19,276 (76.3%). The median difference between these dates was 84 days (IQR 32, 147);
- startART was in a year before approxART for 5,987 (23.7%). Almost half (2,698 [45.1%]) of these individuals had approxART in 1997 or 1998 (number of ARV drugs was not reported before 1997 and may have been incomplete in 1997). Of the latter, 68.6% had startART in 1995, 1996 or 1997 (Figure 8.4).

Individuals with startART in a year before approxART accounted for the majority of individuals with approxART in 1997 and 1998 (Figure 8.4) suggesting that most of these individuals probably started therapy in previous years but did not have that information captured by SOPHID (as assumed).

There were 14,875 individuals who had no reported startART and who had never been reported to be on ART (no approxART). There were 7,197 individuals with approxART who had no reported startART but only 318 who had not been reported to have received ART but who had startART reported.

Figure 8.3. Flow diagram showing the reporting of approxART and startART and the chronological association between them.

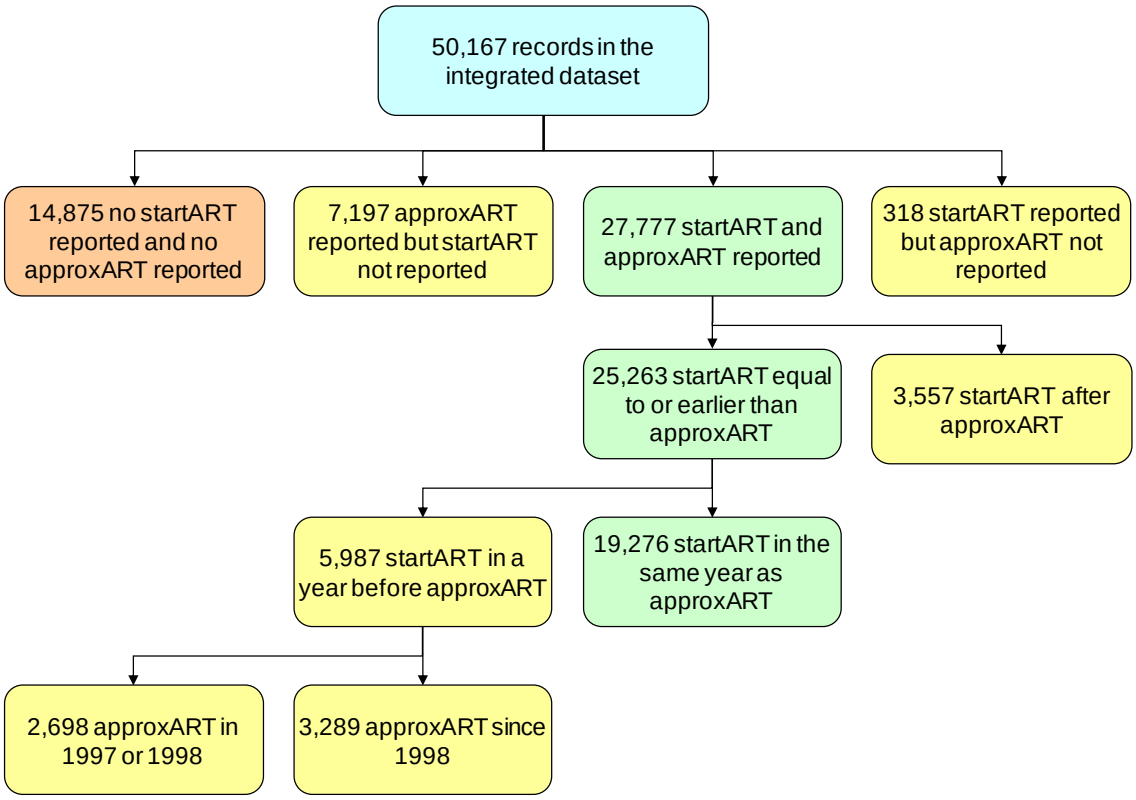
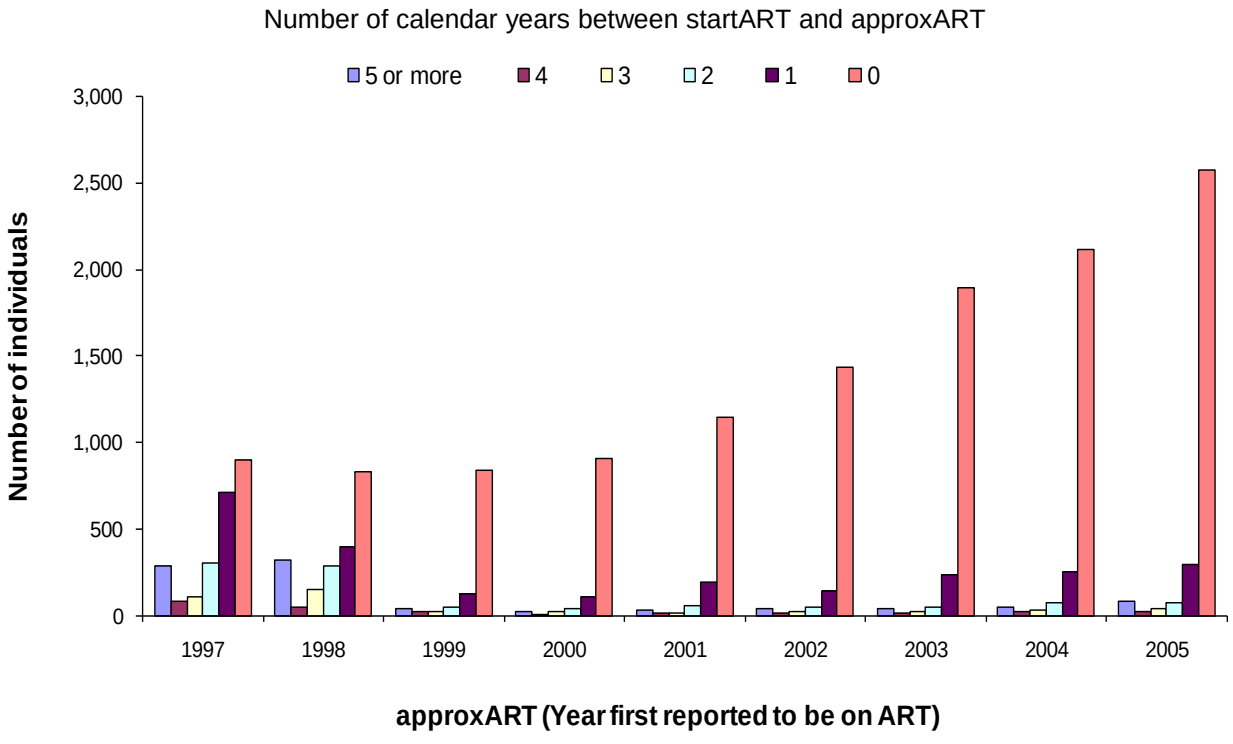


Figure 8.4. Year in which ART was started based on startART stratified by the year of approxART.



8.4.2 Describing the time between CD4 counts and startART

There were 15,567 individuals with approxART after 1998, with startART earlier and in the same year as approxART, and with at least one CD4 cell count earlier than approxART and in the same year.

There were 23,492 individuals who had a CD4 count before startART (where startART was not after approxART), among whom the median time from the last CD4 count before startART to startART was 21 days (IQR 8, 42). Almost a third of these individuals (29.8% [7,010]) started ART more than 30 days after a CD4 count and 7.4% (1,735) started more than 91 days after a CD4 count.

There were 23,555 individuals who had a CD4 count after startART (where startART was not after approxART). The median time from startART to the next CD4 count was 39 days (IQR 21, 92) and 877 individuals had more than 91 days between startART and the next CD4 count. For 18,870 of these individuals who also had CD4 counts before startART, the median time from the last CD4 count before startART to startART was 21 (IQR 8, 41) days. The median time from the first CD4 count after startART to the subsequent CD4 count was 56 days (IQR 29, 91) for 18,361 individuals with subsequent CD4 counts.

There were 14,887 individuals whose CD4 cell count closest to startART was in the same year as both startART and approxART, before approxART and after 1998. The time between this CD4 count and startART varied from 0 days to 352 days (median 14 days [IQR 5, 26]) and there were 12,466 (83.7%) individuals with their closest CD4 count within 30 days of startART.

8.4.3 Evaluating the use of CD4 counts to determine estART among a sample of individuals with startART reported

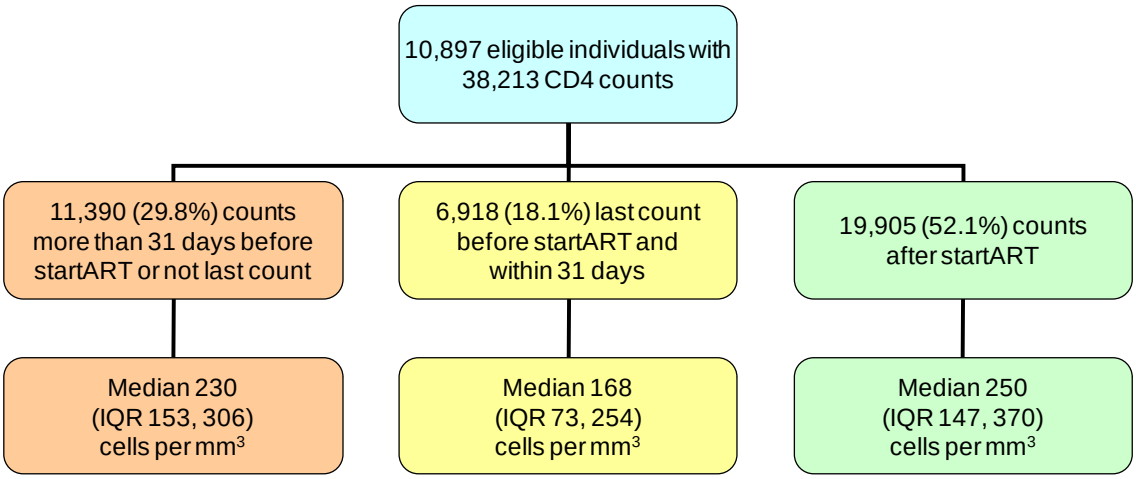
There were 15,567 individuals with approxART after 1998, startART earlier than approxART and in the same year and at least one CD4 cell count measured earlier than approxART and in the same year.

The random sample of 70% used to develop the algorithm consisted of 10,897 of the 15,567 eligible individuals. These individuals had 38,213 CD4 cell counts measured earlier than approxART and in the same year and 8,933 (82.0%) individuals had CD4 counts within 31 days of startART.

8.4.3.1 Evaluating the first time an individual met the recommended criteria for starting ART: CD4 threshold or AIDS

All 10,897 individuals had a CD4 cell count earlier than approxART and in the same year. Of their 38,213 CD4 cell counts, 9,427 (24.7%) were the last before startART and 6,918 (18.1%) of these were within 31 days of startART (Figure 8.5). There were 2,113 individuals with 3,113 CD4 counts within 31 days of an AIDS diagnosis. The median of these was 71 [IQR 23, 160] cells/mm^3.

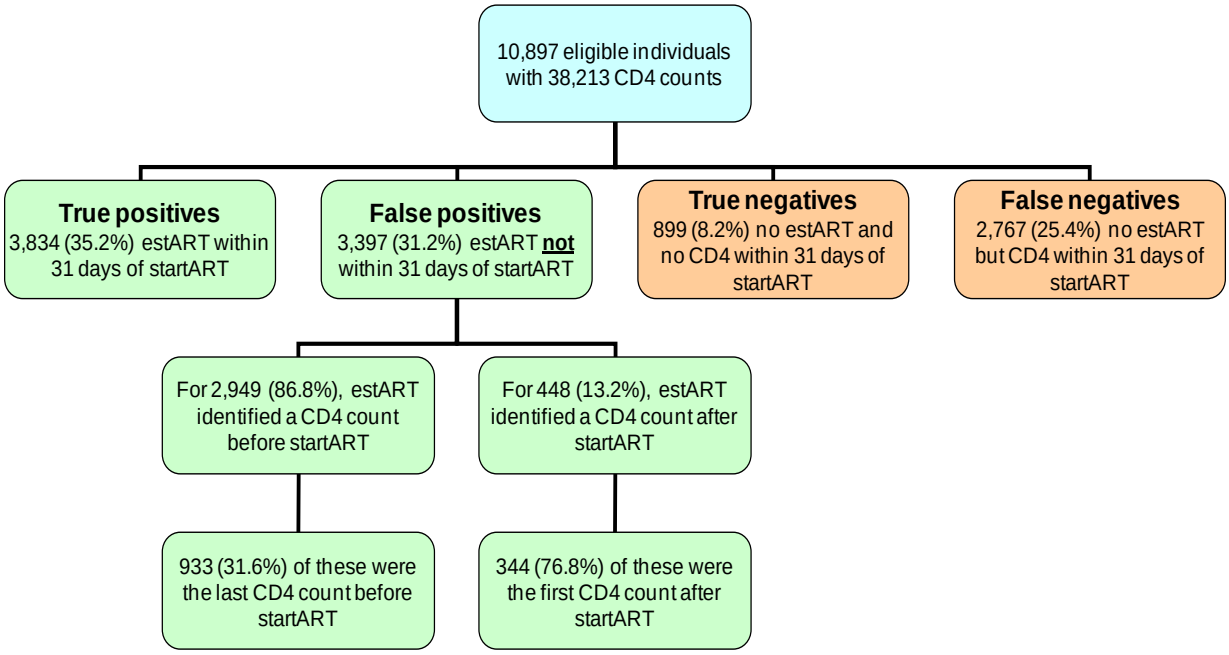
Figure 8.5. Distribution of CD4 counts in relation to startART



There were 526 individuals who started ART before the guidelines changed in 2000. Of these, 444 (84.4%) had CD4 counts below the threshold of 350 cells/mm^3 (of whom, 301 had CD4 counts less than 200 cells/mm^3) and 13 (2.5%) had CD4 counts of 350 cells/mm^3 or more but a previous AIDS diagnosis. Among the 10,371 individuals who started ART after 1999, 6,422 (61.9%) had CD4 counts below the threshold of 200 cells/mm^3 and 428 (4.1%) had CD4 counts of 200 cells/mm^3 or more but a previous AIDS diagnosis (a further 2,648 [25.5%] individuals had CD4 counts of 200-350 cells/mm^3 but no previous AIDS diagnosis).

The earliest CD4 cell count at which the individual fulfilled the criteria for starting ART identified estART within 31 days of startART for 3,834 (35.2%) individuals (Figure 8.6). There were 3,397 (31.2%) individuals with 'incorrect' estART, and 3,666 (33.6%) individuals with no estART because they did not have CD4 cell counts that met the criteria. This algorithm correctly identified estART within 31 days of startART with a sensitivity of 40.2% and a specificity of 45.1%. A substantial percentage of false positives identified the last CD4 count before startART or the first CD4 count after startART (Figure 8.6). This occurred most frequently when an individual did not have a CD4 count measured within 31 days of startART.

Figure 8.6. Distribution of estART in relation to startART



The analysis was also repeated considering CD4 cell counts in the sample that were the only CD4 cell count before approxART to define estART in addition to those defined by the guidelines. There were 3,549 individuals with only one CD4 cell count before approxART, of which the median was 113 cells/mm^3 [IQR 39, 210]. The median difference between estART and startART for these individuals was 19 days [IQR 9, 32] and 73.6% (2,613/3,549) had estART within 31 days of startART. This percentage was only slightly higher among those who fit the criteria for starting ART (76.1% [2,039/2,678]) than among those who didn't (65.9% [574/871]), which was statistically significant ($p<0.01$). Including the only CD4 cell count before approxART in the algorithm along with those defined by the guidelines resulted in estART correctly identified within 31 days of startART for 40.2% (4,381) of individuals, with a sensitivity of 46.0% and a specificity of 36.0%.

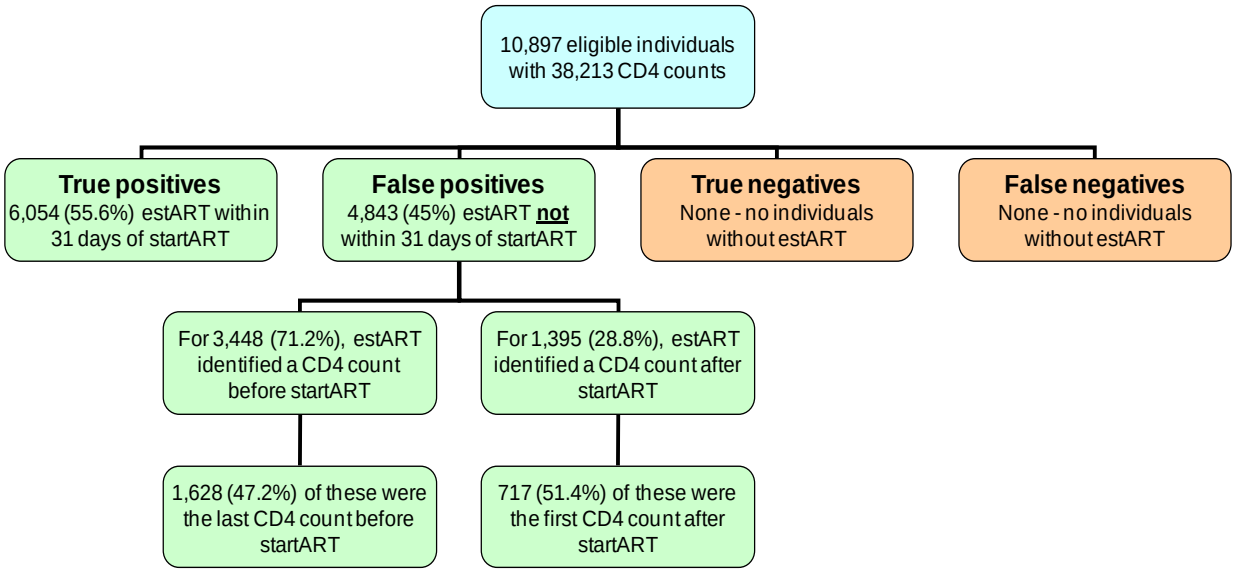
To account for the flexibility in the guidelines, this analysis was repeated using: 1) thresholds of 400 cells/mm^3, 350 cells/mm^3 and 250 cells/mm^3; 2) thresholds of 450 cells/mm^3, 350 cells/mm^3 and 300 cells/mm^3 respectively for people with approxART in 1999, 2000 and 2001 onwards. AIDS was still considered as a prompt for ART initiation. The number of true positives, false positives, true negatives and false negatives were: 1) 4,134, 4,496, 601, 1,666 and 2) 4,423, 5,252, 385, 1,017 respectively. The sensitivities and specificities of these algorithms were 1) 43.4% and 30.2%, and 2) 44.5% and 19.3% respectively.

Of the overall 27,270 individuals with approxART after 1998, 2,503 (9.2%) individuals could not be categorised using this algorithm because they did not have a CD4 count before approxART and in the same year as approxART.

8.4.3.2 Evaluating the earliest CD4 nadir

The 10,897 individuals used to evaluate this algorithm were the same as those included in the analysis above. The median difference between estART (date of CD4 nadir) and startART was 16 days [IQR 0, 40]). The nadir CD4 count correctly identified estART within 31 days of startART for 6,054 (55.6%) individuals (Figure 8.7). By definition, there were no individuals without estART (as all individuals had at least one CD4 cell count and hence a nadir value) and no negative results from the algorithm; therefore, there were no true negative results. A substantial percentage of false positives identified the last CD4 count before startART or the first CD4 count after startART (Figure 8.7) and most frequently when an individual did not have a CD4 count measured within 31 days of startART.

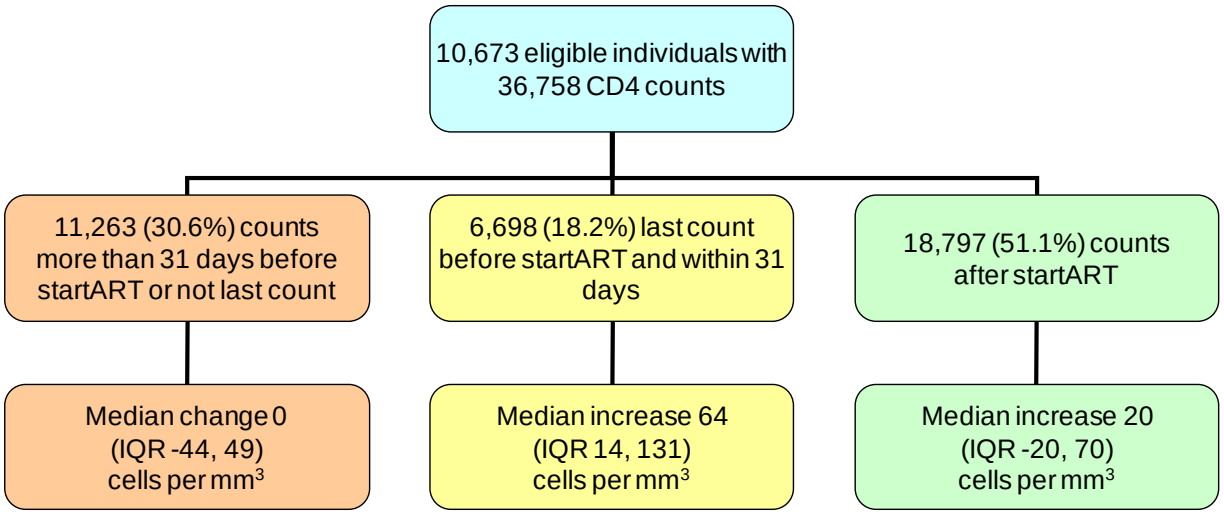
Figure 8.7. Distribution of estART in relation to startART



8.4.3.3 Evaluating a single increase in CD4 counts

There were 10,673 individuals (97.9% of the random sample) who had more than one CD4 cell count measured before or on the same date as approxART and in the same year (Figure 8.8). These individuals contributed a total of 36,758 pairs of consecutive CD4 counts, of which 6,698 (18.2%) spanned startART (where the first was measured within 31 days of startART). However, there was substantial overlap in the range of increases in CD4 counts between the three categories, indicating probable misclassification using this algorithm.

Figure 8.8. Distribution of CD4 counts in relation to startART



Using this algorithm, the sensitivity decreased and the specificity increased as the cut-off increased. Overall, however, the percentage correctly classified fell from 55.0%, with a cut-off of 20 cells/mm^3, to 21.6% with a cut-off of 300 cells/mm^3. There was a peak in the likelihood ratios of 0.860 at 150 cells/mm^3 with a sensitivity of 23.3% and specificity of 72.9% (highlighted in Table 8.1).

Table 8.1. Summary table for a single increase in CD4 counts

Cut-off for change in CD4 counts / cells/mm^3	Number of positives[1]	Number of true positives[2]	Number of negatives[3]	Number of true negatives[4]	Sensitivity	Specificity	Correctly Classified	Likelihood ratio
0	10,190	5,390	483	257	62%	13%	53%	0.714
20	9,739	5,479	934	330	63%	19%	55%	0.788
40	8,971	5,147	1702	396	60%	26%	53%	0.810
60	7,895	4,619	2778	465	54%	36%	50%	0.842
80	6,791	3,975	3,882	538	46%	45%	46%	0.837
90	6,243	3,646	4,430	645	42%	49%	44%	0.835
100	5,686	3,343	4,987	743	39%	54%	42%	0.836
110	5,113	3,008	5,560	840	35%	58%	39%	0.839
120	4,631	2,717	6,042	918	31%	62%	37%	0.837
130	4,200	2,461	6,473	1,008	29%	66%	36%	0.842
140	3,787	2,218	6,886	1,095	26%	70%	34%	0.851
150	3,400	2,010	7,273	1,193	23%	73%	33%	0.860
160	3,044	1,790	7,629	1,273	21%	76%	31%	0.849
170	2,728	1,592	7,945	1,349	18%	78%	30%	0.825
180	2,448	1,421	8225	1,424	16%	80%	29%	0.815
190	2,166	1,252	8,507	1,488	15%	82%	27%	0.822
200	1,952	1,117	8,721	1,542	13%	84%	27%	0.807
210	1,750	1,006	8,923	1,584	12%	85%	26%	0.785
220	1,575	906	9,098	1,628	10%	87%	25%	0.790
230	1,420	799	9,253	1,680	9.3%	88%	24%	0.761
240	1,277	718	9,396	1,713	8.3%	89%	24%	0.764
250	1,155	641	9,518	1,737	7.4%	90%	23%	0.739
260	1,042	575	9,631	1,769	6.7%	91%	23%	0.727
270	938	515	9,735	1,792	6.0%	92%	22%	0.708
280	854	467	9,819	1,818	5.4%	92%	22%	0.717
290	781	432	9,892	1,835	5.0%	93%	22%	0.745
300	711	389	9,962	1,853	4.5%	94%	22%	0.741

[1] positives were where estART was estimated using the algorithm

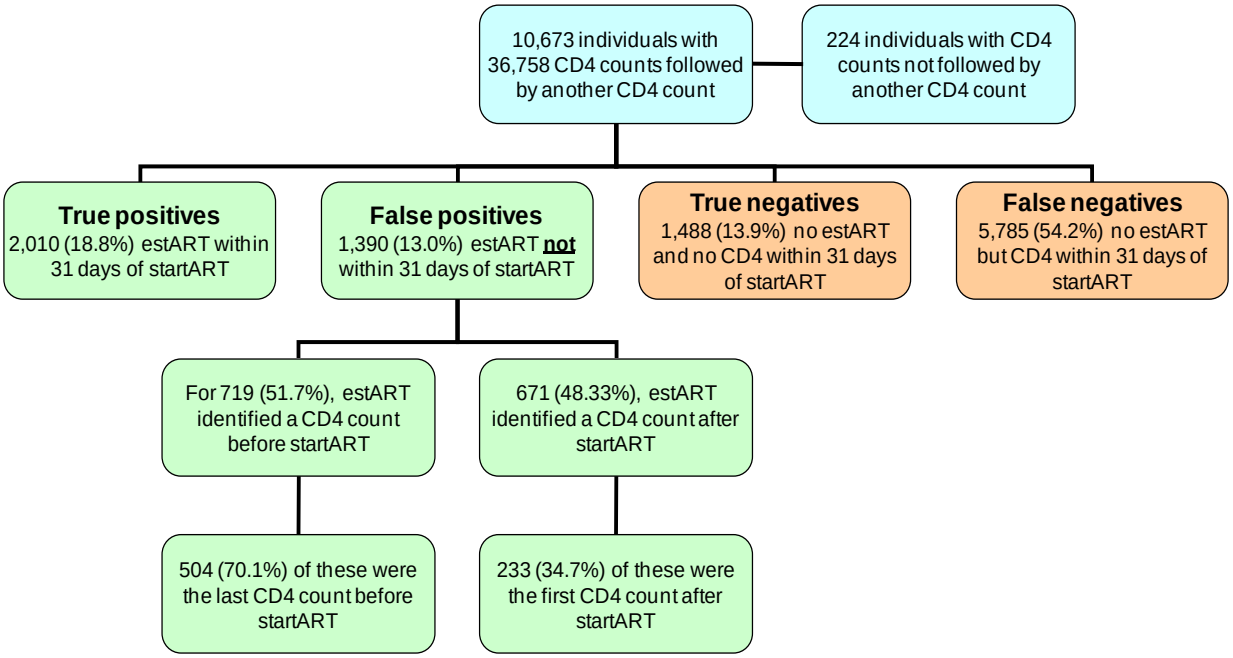
[2] true positives were where estART was within 31 days of startART

[3] negatives were where no estART could be estimated

[4] true negatives were where no estART could be estimated and no CD4 counts measured within 31 days of startART.

Using the cut-off of 150 cells/mm^3, there were 2,010 (18.8%) individuals with estART within 31 days of startART, 1,390 (13.0%) individuals with 'incorrect' estART, and 7,273 (68.1%) individuals with no estART because they did not have CD4 counts that met the criteria (Figure 8.9). There were 224 individuals with no estART because they had CD4 cell counts before approxART and in the same year as approxART but which weren't followed by another CD4 count. The median difference between estART and startART was 7 days [IQR -21, 27]). Of the 1,390 individuals with 'incorrect' estART, 719 (51.7%) had estART more than 31 days before startART with a median difference of 53 [IQR 40, 85] days and 671 (48.3%) had estART more than 31 days after startART with a median difference of 80 [IQR 50, 121] days. Yet estART identified the last CD4 count before startART or the first CD4 count after startART for a substantial percentage of false positives (Figure 9).

Figure 8.9. Distribution of estART in relation to startART



Of the overall 27,270 individuals with approxART after 1998, 3,203 (11.7%) could not be categorised using this algorithm because they did not have a CD4 count before approxART and in the same year as approxART that was followed by another CD4 count.

8.4.3.4 Evaluating other algorithms

The structure of the results from evaluation of algorithms 4-8 was very similar to the single increase in CD4 cell counts and therefore figures and tables specific to the other algorithms are presented in Appendix C. Table 8.2 below summarises the results for all algorithms investigated.

The number of individuals and CD4 cell counts included in the analysis decreased and the overall number of individuals who could not be categorised by the algorithm increased as the number of CD4 cell counts required for the algorithm increased. For example, determination of whether CD4 counts meet the recommended criteria for starting ART and determination of CD4 nadir require only one CD4 cell count before approxART and in the same year, and have the lowest overall percentages (9.2%) who could not be categorised. In contrast, the preceding decrease and subsequent increase in smoothed CD4 counts and the preceding negative slope and subsequent positive slope in CD4 counts require two CD4 cell counts before and two CD4 cell counts after each CD4 cell count used in the algorithm. An overall percentage of 40% of individuals could not be categorised using these algorithms.

The percentage of true positives and true negatives varied substantially such that the percentage correctly classified, with this selection of cut-offs, varied

from 30%, when using the preceding decrease and subsequent increase in smoothed CD4 cell counts and a cut-off of 130 cells/mm^3 per year, to 56% when using the CD4 nadir. However, the positive predictive value is a more appropriate indicator of the usefulness of an algorithm for determining a reliable estART (and therefore increasing the proportion of individuals with this information for analysis). Among these algorithms with this selection of cut-offs, the positive predictive value varied from 46%, when using the preceding decrease and subsequent increase in CD4 cell counts, to 68% when using the positive slope in CD4 cell counts. The absolute percentage of true positives must also be considered as an indicator of the usefulness of an algorithm – this varied from 3.6%, when using the preceding decrease and subsequent increase in smoothed CD4 cell counts, to 56% when using the CD4 nadir.

Table 8.2. Summarised results for all algorithms

	Algorithm							
	1) First CD4 to meet criteria for starting ART	**2) CD4 nadir**	**3) Single CD4 increase**	**4) Increase in smoothed CD4 counts**	**5) Preceding decrease and subsequent increase in CD4 counts**	**6) Preceding decrease and subsequent increase in <u>smoothed</u> CD4 counts**	**7) Positive slope in CD4 counts**	**8) Preceding negative slope and subsequent positive slope in CD4 counts**
Cut-off	350 cells/mm^3/200 cells/mm^3/AIDS	n/a	150 cells/mm^3	110 cells/mm^3	*150 cells/mm^3*	*130 cells/mm^3*	*800 cells/mm^3 per year*	*800 cells/mm^3 per year*
Number of individuals included in evaluation, n (%)	10,897 (100%)	10,897 (100%)	10,673 (97%)	8,945 (82%)	9,284 (85%)	7,513 (69%)	9,284 (85%)	7,513 (69%)
Number of counts included in evaluation, n	38,213	38,213	36,758	28,933	30,401	23,499	30,401	23,499
Median time from estART to startART (IQR) / days	24 (7, 49)	16 (0, 40)	7 (-21, 27)	0 (-35, 18)	14 (-26, 43)	5 (-21, 29)	-14 (-30, 2)	15 (-2, 40)
True positives, n (%)	3,834 (40%)	6,054 (56%)	2,010 (18%)	1,190 (11%)	1,409 (13%)	394 (4%)	1,508 (14%)	1,001 (9%)
False positives, n (%)	3,397 (31%)	4,843 (44%)	1,390 (13%)	823 (8%)	1,680 (15%)	313 (3%)	701 (6%)	880 (8%)
True negatives, n (%)	899 (8%)	n/a	1,488 (14%)	2,212 (20%)	1,994 (18%)	2,890 (27%)	2,351 (22%)	2,783 (26%)
False negatives, n (%)	2,767 (25%)	n/a	5,785 (53%)	4,720 (43%)	2,043 (39%)	3,916 (36%)	4,724 (43%)	2,849 (26%)
Correctly classified, n (%)	4,733 (43%)	6,054 (56%)	3,498 (32%)	3,402 (31%)	3,403 (31%)	3,284 (30%)	3,859 (35%)	3,784 (35%)
Positive predictive value (%)	53%	56%	59%	59%	46%	56%	68%	53%
Overall number that could not be categorised, n (%)	2,503 (9%)	2,503 (9%)	3,203 (12%)	7,630 (28%)	6,896 (25%)	10,897 (40%)	6,896 (25%)	10,897 (40%)

8.4.3.5 Selecting the most appropriate algorithm

Having investigated a number of possible algorithms there were two approaches to achieve the aim of estimating a reasonably accurate estART for individuals known to have started ART in any particular year (known approxART), but in whom startART was unknown:

1. Use a single algorithm and select one with an acceptable sensitivity and specificity.

   The use of recommended criteria for starting ART and the CD4 nadir were not considered to have sufficient specificity. The algorithm using the slope through three consecutive CD4 cell counts and the cut-off of 800 cells/mm^3 per year had high sensitivity at specificities of close to 80% and 90% (Table 8.3). Additionally, the number of individuals who could not be categorised using this algorithm (6,896 [25%]) was not as great as that for the algorithm requiring a preceding negative slope and subsequent positive slope (10,897 [40%]), which had similar sensitivity. Therefore, this algorithm with the cut-off of 800 cells/mm^3 per year was selected as the most appropriate to achieve the aim of estimating estART for as many individuals as possible with reasonable accuracy (sensitivity of 23% and specificity of 91%). This correctly identified estART for 1,508 individuals, incorrectly identified estART for 701 individuals and left 8,688 individuals with approxART as the best estimate (1,613 because they did not have a CD4 count before approxART and in the same year that was both preceded and followed by other CD4 counts).

Table 8.3. Comparative sensitivities, specificities and percentages uncategorised for different algorithms (two sets of thresholds that produced specificities closest to 90% [panel 1] and 80% [panel 2] were used for comparison of sensitivity and percentages uncategorised])

Algorithm number and description	Sensitivity	Specificity	Overall percentage uncategorised
Using thresholds producing specificities closest to 90% for comparison			
3) Single increase in CD4 counts	7.4%	90%	12%
4) Increase in smoothed CD4 counts	14%	90%	28%
5) Decrease then increase in CD4 counts	8.0%	89%	25%
6) Decrease then increase in smoothed CD4 counts	18%	89%	40%
7) Positive slope through CD4 counts	23%	91%	25%
8) Negative then positive slope through CD4 counts	24%	90%	40%
Using thresholds producing specificities closest to 80% for comparison			
3) Single increase in CD4 counts	16%	80%	12%
4) Increase in smoothed CD4 counts	26%	80%	28%
5) Decrease then increase in CD4 counts	18%	79%	25%
6) Decrease then increase in smoothed CD4 counts	31%	79%	40%
7) Positive slope through CD4 counts	37%	79%	25%
8) Negative then positive slope through CD4 counts	33%	78%	40%

2. Use the best algorithm for individuals with sufficient CD4 cell counts followed by an inferior algorithm that can be applied to additional individuals and using an increased cut-off to maintain specificity with this algorithm.

    a. The first algorithm selected used the slope through three consecutive CD4 counts and a cut-off of 800 cells/mm^3 per year. This had a specificity of 91%, sensitivity of 23, and 25% remained uncategorised.

    b. The next algorithm selected used a single increase in CD4 cell counts only for the 25% (6,896) of the 27,270 individuals with an approxART after 1998 who did not have enough CD4 counts to calculate a slope. A cut-off of 250 cells/mm^3 was selected because this had a similarly high specificity of 90% (although the sensitivity was only 7.4%).

The sensitivity of these two combined algorithms was slightly lower than that of the slope used alone (20% versus 23%) although the specificity remained at 90%. However, the combined algorithms performed markedly better than the single increase in CD4 counts alone. There were 1,389 additional individuals categorised including 178 true positives and 99 false positives compared to using the slope alone. Of the sample of 10,897 individuals, there remained 2.1% (compared to 15% using the slope alone) where the combined algorithm could not estimate an estART and the best estimate of startART was approxART. There remained 3,203 (12%) individuals who could not be categorised using this combined algorithm (compared to 6,896 [25%] using the slope alone) because of insufficient CD4 counts.

The investigation of a number of possible algorithms and combinations of algorithms indicated that the 'best option' for determining an accurate estART for individuals without startART reported was to use the combination of algorithms described above. This provided the best sensitivity with reasonable specificity compared to the other algorithms while also minimising the number of individuals who could not be categorised because of insufficient CD4 counts.

8.4.4 Validation of the 'combined' algorithm

The 'combined' algorithm was validated on the 30% (4,670) of individuals who were not included in the 70% sample used above. There were 4,583 (98%) of these individuals who had at least two CD4 cell counts so that the algorithm could be applied. The combined algorithm identified estART for 1,049 (22%) of

all the individuals. estART was a median of 83 [IQR 42, 133] days before approxART with a range of 0-343 days. There were 716 (15%) individuals with estART within 31 days of startART, 333 (7%) individuals with 'incorrect' estART, and 3,534 (78%) individuals with no estART.

There were 130 (39%) individuals with an 'incorrect' estART that was more than 31 days before startART with a median difference of 60 [IQR 41, 83] days, and 203 (59%) individuals where estART was more than 31 days after startART with a median difference of 57 [IQR 42, 84] days. Yet estART identified the last CD4 count before startART for 73 (56%) of these individuals and the first CD4 count after startART for 67 (33%) of these individuals respectively.

8.4.5 Applying the algorithm to individuals without startART

There were 7,197 individuals with approxART who had no reported startART, of whom 4,862 had approxART after 1998. Of these, 4,162 individuals had CD4 cell counts earlier but in the same year as approxART and 3,889 (54%) had sufficient CD4 cell counts to determine single increases in CD4 cell counts for the algorithm.

The algorithm identified estART for 578 (15%) of the 3,889 individuals. EstART was a median of 91 [IQR 35, 153] days before approxART with a range of 0-344 days. However, approxART remained the most reliable estimate of the date of starting ART for the majority of patients.

8.4.6 Combination of reported and estimated ART start dates

There were 27,270 individuals with approxART after 1998 of whom 22,408 (82.2%) had a startART reported and a further 578 (2.1%) had an estART determined by the algorithm. A remaining 4,284 (15.7%) individuals had approxART as the best estimate of startART. Of these:

- 339 (7.9%) had no reported CD4 count prior to approxART
- 361 (8.4%) had a CD4 count reported prior to approxART but not within the same calendar year
- 273 (6.4%) had a CD4 count reported prior to approxART and within the same year but did not have sufficient CD4 counts to determine the algorithm
- 3,311 (77.3%) had a CD4 count reported prior to approxART and within the same year and sufficient CD4 counts to determine the algorithm but the algorithm failed to identify an estimated ART start date.

The conclusion was that the algorithm was not sensitive and specific enough to accurately identify startART for a sufficient number of individuals to justify the introduction of the additional methodology and the additional false positive results.

8.4.7 Sensitivity analysis – evaluation of the combined algorithm with startART in the 70% sample considering CD4 counts in the same year as approxART and the year before approxART

Inclusion of CD4 counts in the year before approxART increased slightly the number of individuals that could be included in the sensitivity analysis (Table 8.4). The number correctly classified, sensitivity and specificity decreased (although negligibly) but the reduction in the overall percentage that could not be classified (from 12% to 10%) was insufficient to consider use of this version of the algorithm.

Table 8.4. Summarised results for the combined algorithm considering CD4 cell counts in the same year as approxART and the year before approxART compared to the combined algorithm considering only CD4 cell counts in the same year as approxART.

	Only CD4 counts in same year as approxART	**CD4 counts in same year and year before approxART**
Number of individuals included in evaluation n (%)	10,673 (98%)	10,753 (98%)
Number of counts included in evaluation / n	36,758	40,912
Median time from estART to startART (IQR) / days	-10 (-28, 10)	-7 (-28, 13)
True positives / n (%)	1,686 (16%)	1,680 (15%)
False positives / n (%)	800 (7%)	863 (8%)
True negatives / n (%)	1,839 (17%)	1,831 (17%)
False negatives / n (%)	6,348 (59%)	6,299 (57%)
Correctly classified / n (%)	3,525 (33%)	3,511 (33%)
Overall number that could not be categorised / n (%)	3,203 (12%)	2,708 (10%)

8.5 Discussion

This chapter aimed to produce a robust estimate of the date of starting ART for individuals missing this crucial piece of information. This information is regularly used to assess incidence rates of events or CD4 trajectories before and after starting ART, for assessing when people start therapy and as a baseline in studies of response to ART. The HPA has now started to collect dates of starting ART to monitor important public health outputs and indicators. This is a prime example of surveillance developing in association with public health data needs – in this case, together with the performance monitoring and quality assurance aspects of the developing commissioning role in the NHS.

A surprising conclusion from these analyses was that the algorithm based upon treatment guidelines was not even reasonably accurate in determining the date of starting ART even when various sensitivity analyses were considered. The majority of the false positive results identified CD4 cell counts more than 31 days before the date of starting ART indicating delays between the recommended time to start ART and the actual date of initiation. Similar findings have been published raising public health interest for further investigation[325;495;506;507]. In the UK, during the period analysed, there were no guidelines for the timeliness of ART initiation. Subsequent guidelines, which may partly reflect previous practice, state that clinicians should assess patients' readiness to start therapy and allow them to make decisions about their therapy. Yet, that ART should be started immediately for patients with certain AIDS diagnoses, co-morbidities or very low CD4 counts and within two weeks

for other patients with AIDS, serious bacterial infections or CD4 counts less than 200 cells per mm^3 [128]. Therefore, it is reasonable to expect most patients to have started ART within 31 days after becoming eligible by CD4 count or by AIDS diagnosis, as in this analysis. Hence, reasons for substantially delayed ART initiation should be determined and addressed to reduce this period when patients are at high risk of AIDS and death. Additionally, a third of individuals starting ART after 1999 did not meet the eligibility criteria, which should also be investigated further as it could reflect inappropriate advice provided to patients.

The analyses showed that almost a third of CD4 nadirs occurred more than 31 days before starting ART and 13% occurred more than 31 days after starting ART. This finding was not useful for the determination of an estimated date of starting ART. However, it may be useful for researchers analysing longitudinal data because it may indicate that additional validation is necessary before the use of a single CD4 count as a baseline measure of immunosuppression prior to ART[233;234]. This was recognised in UK treatment guidelines in 2008[123] (but not re-stated in 2012 guidance[128]). Among the sample of 9,488 where baseline CD4 was the one closest to and within 31 days of startART, the median difference between the CD4 nadir and the baseline CD4 count was 0 (IQR 0, 41) cell per mm^3 but the difference was greater than 50 cells/mm^3 for 22.6% (n=2,145) and greater than 100 cells/mm^3 for 11.8% (n=1,124). These results also suggest further analysis of factors associated with CD4 counts declines after the start of ART among 13% of individuals. However, this would also require assessment of the effects of natural and measurement variability, further validation of the data and consideration of statistical techniques[494].

CD4 cell counts change due to pregnancy among non-HIV-infected women (Section 1.3.2) but studies show no significant change among HIV-infected women[508-510]. Therefore, there is no difference in the recommendation for therapy by CD4 cell count for pregnant and non-pregnant women according to guidelines[135] and CD4 counts during pregnancy can be considered to accurately reflect levels of immunosuppression and CD4 nadirs for use in these algorithms.

In this study, the median increase between two consecutive CD4 cell counts that spanned the start of ART was 64 cells/mm^3 and the median subsequent increase was 20 cells/mm^3. Less than a third of individuals had single CD4 count increases that exceeded the cut-off of 150 cells/mm^3 and more than two-fifths of those did not occur within 31 days of starting ART. Similar increases in CD4 cell counts of a median 130 cells per mm^3 within six months of starting ART and 176 cells per mm^3 within a year are reported in the literature[511]. However, there are no published data on the proportion of individuals who do not experience an increase or the proportion that experience similar increases before they start ART. The only relevant information published comes from cohort studies on treatment-naïve individuals, which indicate that a quarter do not experience increases greater than 24 cells/mm^3 [512] or 51 cells/mm^3 [498] after six months. The only range published was a 12 months increase, among individuals who controlled their viraemia, which showed that decreases of up to 200 cells/mm^3 occurred[364].

The algorithm using a preceding decrease and subsequent increase in CD4 counts performed worse than the one using a single increase in CD4 counts. The sensitivities and specificities were similar but a higher percentage of individuals did not have sufficient data to be categorised. Only a fifth of individuals exceeded the cut-off within 31 days of starting ART, whilst more than half of those occurred after more than 31 days. The literature has focused on quantifying decreases in CD4 counts before starting ART and increases in CD4 counts after starting ART but overlooked quantification of the combination of the two and the proportions that do/do not experience declines before and increases after, which could help describe the progression of HIV infection[485].

The algorithm based on a positive slope in CD4 counts performed best but was still not sufficient for the estimated date of starting ART to be used in subsequent analyses. It was clear that many individuals did not have a substantial positive slope at the time of starting ART but that many individuals experienced such increases even before ART was started. False positive results accounted for 29% of all positives even when the cut-off rate was 1,600 cells/mm^3 per year (double the cut-off selected). High rates of increase shortly after starting ART are known, with published rates of of 11 cells/mm^3 per week during the first 8 weeks among all treatment-naïve individuals[513] and 30 cells/mm^3 per month before the fourth month[514] and 97 cells/mm^3 per month during the first month[364] among individuals who controlled their viraemia. Furthermore, the data showed that a fair proportion of the false positives were the last CD4 count before starting ART or the first CD4 count after starting ART even though these did not occur within 31 days of starting ART.

Some ART regimens may be more effective than others at increasing CD4 counts* and this may have resulted in a differential bias of identifying estART for individuals on those regimens[515;516]. This would possibly bias the results of subsequent analyses based on estART towards inclusion of individuals on those regimens. However, ART regimens are not available in the integrated dataset to quantify this potential bias as they are not reported to surveillance and other variations in rates of increase of CD4 counts are much more substantial. Therefore, this was not believed to be among the most important limitations of these analyses.

It is possible that a proportion of the false positives were due to reporting errors in the date of starting ART or due to CD4 measurement errors. Studies to develop these investigations would need to examine a sample of false positives for verification. Even when slopes both before and after each CD4 cell count were considered or a smoothing mechanism was employed, the specificity was even lower and did not exclude the false positives.

Some individuals did not have CD4 cell counts within 31 days of starting ART but the algorithms did identify many of the last CD4 counts before starting ART or the first CD4 counts after the date of starting ART for these individuals. A recent CD4 count should be available for all individuals to inform the initiation of ART and the lack of this for all individuals could be investigated further for

* Of the order of 10-15% higher CD4 count after 48 weeks for ritonavir-boosted PI regimens than PI or NNRTI regimens and 24% higher than for NRTI-only regimens.

public health interest although the possibility of unlinked records would have to be excluded.

8.6 Conclusion

Overall, this analysis indicates that the assumption that most individuals will experience a CD4 cell count below the threshold recommended by guidelines or AIDS at the time of starting ART, and not before, is not valid. Neither is the assumption that most individuals will experience a CD4 nadir at the time of starting ART and not before. Furthermore, individuals may experience substantial increases in CD4 counts, positive slopes in CD4 counts or even preceding decreases/subsequent increases in CD4 counts more than a month before the time of starting ART.

While the outcome of these analyses did not determine an algorithm that could be used to estimate dates of starting ART, the results were very informative about CD4 cell count changes around the time of starting ART. Additionally, for individuals who only have an approximate date of starting ART available for analyses, these results will still inform the allocation of patient follow-up and CD4 counts (as prior to or after starting ART) and the identification of CD4 counts that are highly likely to reflect levels of immunosuppression just prior to starting ART. However, the methods used for censoring patient follow-up and assigning baseline CD4 counts prior to the estimated date of starting ART will have to be evaluated. Finally, these analyses also provide methodological options for detecting potential clerical errors that could be used to improve surveillance data.

Chapter 9. Rates of change of CD4 cell counts after starting ART

9.1 Introduction

Effective ART reduces the risk of morbidity and mortality among HIV-infected individuals by reducing the viral load and increasing the CD4 cell count. Current CD4 counts remain a predictor of morbidity and mortality among HIV-infected individuals on ART and so the rate at which they increase after starting ART is clinically important[360;517-519]. The literature shows that the immunological response to ART is biphasic but reasonably linear in each phase with CD4 counts increasing rapidly during the first 3-18 weeks of ART, followed by slower increases[364;513;514;520-525]. CD4 counts may return to levels similar to those of non-HIV-infected individuals if effective ART can be maintained for a sufficient length of time[526]. Yet, some individuals do not achieve or maintain virological suppression due to poor adherence or drug resistance[527]. CD4 cell count increases are rare without at least transient or partial virological suppression[528]. However, CD4 cell counts may decline for some individuals on ART even if viraemia is controlled (indicating adherence to ART and no resistance) [364;499;503].

Analyses of rates of change of CD4 cell counts after starting ART have helped to further understand responses to ART and determine rates of increase and maximum CD4 cell counts that may be aimed for and achieved [364;498;513;522;526;529;530]. The integrated dataset was therefore used in this chapter to provide population-level information on rates of change of CD4 cell counts after starting ART in E,W&NI and associated factors. Baseline factors

associated with slower rates may indicate unmet need among population groups that could be addressed to improve public health outcomes.

9.2 Aims

a) To use the integrated dataset to investigate baseline factors that were associated with rates of change of CD4 counts:
    a. During the first three months after starting ART;
    b. After the first three months after starting ART.

b) To determine the length of time that baseline CD4 cell counts were associated with rates of increase in CD4 cell counts after starting ART.

9.3 Methods

The time of starting ART used as the baseline in this analysis was taken as: 1) startART where there was no evidence of previous ARV use; or 2) estART, where there was no startART but evidence of ARV use, or evidence of ARV use before startART. Records from the integrated dataset were included in analyses for individuals starting ART after 1998 and with no evidence of starting ART before diagnosis in the UK.

Rates of change of CD4 cell counts for each individual were analysed from the time of starting ART (baseline). To reduce bias in the baseline CD4 cell count due to natural and measurement variability, the mean of the last two CD4 cell counts within the three months before startART or estART were used unless only one was available. Individuals without a CD4 cell count during the three months before starting ART were excluded from the analysis. EstART was

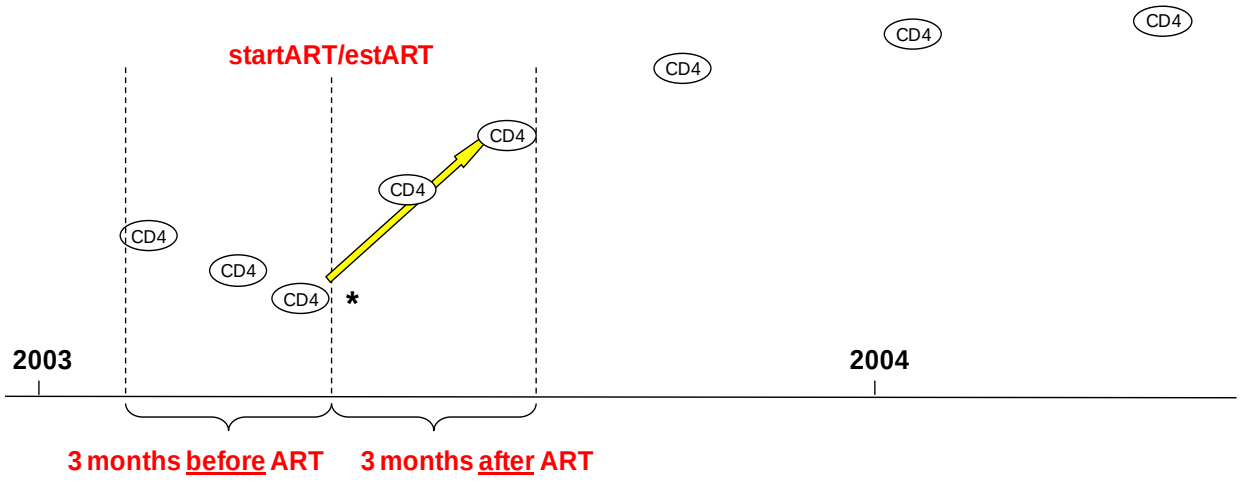
included in the main analysis (but excluded from the sensitivity analysis) to maximise the number of individuals included in the analysis and evaluate the use of estART. Although the algorithm did not produce a good yield or sensitivity, and was therefore not effective for identification of startART among all individuals, good specificity was achieved. Therefore, relatively few false positives would be included in the analysis but also a fair proportion of these were either the last CD4 count before starting ART or the first CD4 count after starting ART even though these did not occur within 31 days of starting ART.

Follow-up time was split into the period during the first three months after starting ART and the subsequent period (Figure 9.1). This was because CD4 cell count increases are known to be much greater during the first few months than subsequently[498;515;520;523]. Additionally, CD4 cell count increases during the first three months are likely to be substantially affected by error in the estimation of estART due to the nature of the integrated dataset (Figure 9.2). This is not applicable to other studies of longitudinal CD4 counts where the date of starting ART is known.

Linear mixed-effects models were used for analysis with overall mean CD4 cell counts considered to be linear over time but allowing for a random intercept and random slope at the individual level. Random intercepts allowed for each individual's vertical shift from overall means and random slopes allowed for each individual's deviation in linear rate of CD4 change from overall mean linear rates of CD4 change. The model allowed for different rates of CD4 increase in those with different baseline characteristics through the inclusion of interactions

between each baseline factor and the time from starting ART. Factors analysed in the analyses were: ethnicity, risk group, age group at baseline, year of starting ART, baseline CD4 cell count category, previous AIDS (prior to startART or estART), and whether treatment was started in London or not. These were all, except for region, shown in previous analyses to be associated with CD4 responses to ART (ethnicity[498;513;516]; risk group[531]; age[516], year[498;515;531], CD4 count[515], previous AIDS[531;532]. Region was included for consistency throughout the thesis and to test geographic generalisability.

Figure 9.1. Schematic showing how baseline CD4 cell counts and rates of increase in CD4 cell counts were defined



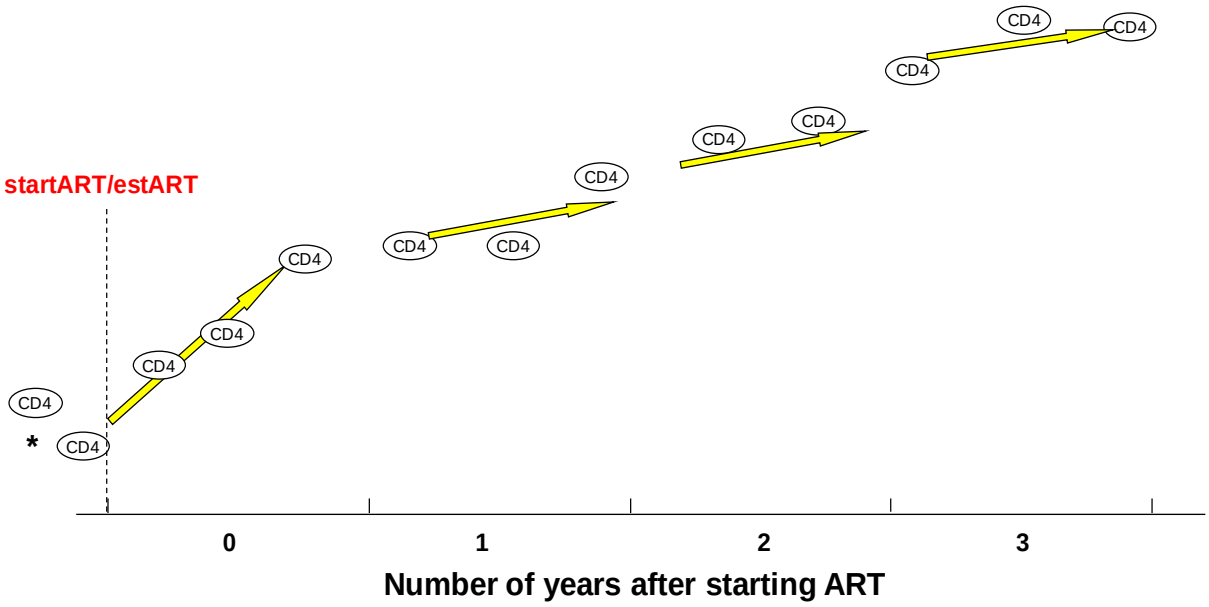
*** Baseline CD4 cell count was the average of the last two CD4 cell counts before startART/estART where available, otherwise the last CD4 cell count**

A number of sensitivity analyses were carried out using i) excluding all individuals who started ART with baseline CD4 counts greater than 349 cells/mm^3; ii) excluding all women; iii) using only the last CD4 count before starting ART to define the baseline; iv) using the mean of all CD4 counts in the

three months before starting ART to define the baseline; v) using only records from individuals who had a specific date of starting ART (startART) reported. The exclusion of individuals with baseline CD4 cell counts greater than 349 cells/mm^3 and, separately, women were to determine the effects of excluding pregnant women who started ART solely to prevent mother-to-infant transmission and may have subsequently interrupted ART. Individuals with one or more CD4 cell counts before starting ART were differentiated, primarily, to investigate whether there were indications of different degrees of bias in the baseline CD4 cell counts arising from regression to the mean. The analysis limited to individuals with startART was to exclude the systematic bias of estART dates being later than the actual dates of starting ART.

In the stratified analysis of long-term follow-up by year to investigate the duration of effect of baseline factors on increases in CD4 cell counts separate linear mixed-effects models were used to determine the association between the baseline factors and the mean rate of increase in CD4 cell counts for each year of follow-up.

Figure 9.2. Schematic showing how rates of increase in CD4 cell counts were calculated for each year after starting ART



*** Baseline CD4 cell count was the average of the last two CD4 cell counts before startART/estART where available, otherwise the last CD4 cell count**

9.4 Results

9.4.1 Rates of increase of CD4 cell counts during the three months after starting ART

There were 25,727 individuals who started ART between 1999 and 2007, of whom 21,203 had a baseline CD4 cell count followed by a median of one subsequent CD4 cell count (range 0, 8) during the three months after starting ART. These 21,203 individuals included in the analysis (those with baseline CD4 cell counts) were not representative of all individuals starting ART (Table 9.1). In multivariable analysis, there were marked differences in proportions included for all factors. Individuals starting ART in more recent years, in London, or with a previous AIDS diagnosis were more likely to be included in the analysis. Individuals were also more likely to be included if they were MSM or middle-aged (40-44 years). Individuals were less likely to have baseline CD4 cell counts, and therefore less likely to be included, if they were Indian\Pakistani\Bangladeshi.

Overall, the mean rate of change of CD4 counts during the first three months of ART was 120 (95% CI 117, 123) cells/mm^3 (Table 9.2). The reference group for analyses was white MSM aged 40-44 years who had a baseline CD4 cell count of 200-249 cells/mm^3 and no previous AIDS diagnosis when they started ART in 1999 in London. This reference group had a rate of change of 149 (95% CI 132, 166) cells/mm^3. All factors except a previous AIDS diagnosis were significantly associated with the initial rate of increase in CD4 counts in multivariable analysis (Table 9.2). There was no significant difference in the initial rate of

increase in CD4 counts between 1999 and 2004 but the rates of increase were a quarter to two-fifths higher between 2005 and 2007 ($p<0.01$). Baseline CD4 counts of 250 cells/mm^3 or more were associated with slower initial rates of increase in CD4 counts ($p<0.01$). In contrast, individuals with baseline CD4 cell counts of 50-99 cells/mm^3 had faster initial rates of increase than the reference group ($p=0.01$). Having a treatment provider in London was associated with a faster initial rate of increase in CD4 counts ($p<0.01$). Individuals aged 15-24 years or 25-29 years had significantly faster rates of increase than individuals aged 40-44 years ($p<0.01$ and $p=0.05$ respectively). IDU and heterosexuals had significantly slower initial rates of increase in CD4 counts compared to MSM ($p<0.01$) except for heterosexual women diagnosed antenatally, who had faster rates of increase ($p<0.01$). Compared to white individuals, black Africans, other black individuals and 'Other Asians' had significantly slower initial rates of increase in CD4 counts ($p<0.01$, $p=0.03$ and $p=0.04$).

Table 9.1. Individuals included in analysis of initial rates of increase in CD4 cell counts (i.e. individuals with baseline CD4 cell counts) as a proportion of all individuals who started ART between 1999 and 2007.

	Number of individuals included in analysis	Proportion of all individuals starting ART (%)	Multivariable odds ratio for inclusion in analysis (95% confidence interval)
Calendar year of starting ART			
1999	1,353	73.3	-
2000	1,495	76.3	**1.19 (1.02 , 1.37)**
2001	1,763	79.3	**1.44 (1.24 , 1.66)**
2002	2,034	73.7	1.06 (0.93 , 1.22)
2003	2,841	83.8	**1.99 (1.73 , 2.29)**
2004	2,875	84.6	**2.13 (1.85 , 2.45)**
2005	3,017	87.7	**2.76 (2.38 , 3.20)**
2006	2,922	85.2	**2.21 (1.92 , 2.55)**
2007	2,903	88.7	**2.98 (2.56 , 3.47)**
Previous AIDS diagnosis			
No	16,624	82.4	-
Yes	4,579	82.4	1.12 (1.04 , 1.22)
Site of care			
Outside London	8,874	82.0	**0.90 (0.84 , 0.96)**
London	12,329	82.7	-
Age group			
15-24	1,487	80.3	**0.76 (0.65 , 0.88)**
25-29	3,295	81.2	**0.85 (0.75 , 0.96)**
30-34	4,940	82.6	0.95 (0.84 , 1.06)
35-39	4,661	83.1	0.98 (0.87 , 1.09)
40-44	3,134	84.0	-
45-49	1,725	82.1	**0.86 (0.75 , 0.99)**
50-54	930	82.7	0.94 (0.79 , 1.13)
>54	1,031	80.7	**0.81 (0.69 , 0.96)**
Risk group			
MSM	8,019	84.2	-
Heterosexual men	4,443	81.3	**0.79 (0.71 , 0.88)**
Heterosexual women	6,588	81.4	**0.80 (0.72 , 0.90)**
Heterosexual women (diagnosed antenatally)	1,577	84.1	0.98 (0.84 , 1.16)
IDU	437	74.6	**0.56 (0.46 , 0.68)**
Recipients of blood products	139	78.5	0.78 (0.54 , 1.13)

Ethnicity			
White	9,300	82.9	-
Black African	9,464	81.9	0.98 (0.88 , 1.09)
Black Caribbean	713	80.9	0.91 (0.75 , 1.09)
Black Other	410	83.8	1.09 (0.84 , 1.41)
Indian\Pakistani\Bangladeshi	270	78.3	**0.76 (0.58 , 0.99)**
Other Asian	341	84.8	1.17 (0.88 , 1.55)
Other/mixed	705	84.9	1.16 (0.95 , 1.42)
Total	21,203	82.4	-

Table 9.2. Results from linear mixed-effects analysis of the association between each factor and the initial rate of increase of CD4 counts during the first three months of ART.

	Initial rate of increase of CD4 count estimated from univariable regression (95% CI) /cells/mm^3	Multivariable regression coefficient (95% CI)	Multivariable p value
Calendar year of starting ART			
1999	102 (91 , 113)	-	-
2000	106 (95 , 116)	-1.4 (-18.3 , 15.4)	0.87
2001	109 (99 , 119)	3.0 (-13.4 , 19.3)	0.72
2002	106 (97 , 115)	2.6 (-13.1 , 18.4)	0.74
2003	110 (103 , 118)	8.5 (-6.4 , 23.5)	0.26
2004	113 (106 , 121)	10.8 (-4.1 , 25.7)	0.16
2005	132 (125 , 139)	**28.8 (14.0 , 43.6)**	**<0.01**
2006	138 (131 , 145)	**35.6 (20.8 , 50.5)**	**<0.01**
2007	141 (133 , 149)	**39.3 (24.0 , 54.6)**	**<0.01**
Baseline CD4 count category (cells/mm^3)			
0-49	121 (113 , 129)	-9.4 (-21.0 , 2.3)	0.12
50-99	151 (141 , 160)	**16.7 (4.3 , 29.0)**	**<0.01**
100-149	145 (136 , 154)	8.0 (-3.9 , 19.9)	0.19
150-199	147 (139 , 155)	4.6 (-6.5 , 15.7)	0.42
200-249	146 (138 , 154)	-	-
250-299	130 (121 , 140)	**-17.4 (-29.3 , -5.4)**	**<0.01**
300-349	123 (111 , 135)	**-26.2 (-40.6 , -11.9)**	**<0.01**
350-499	99 (88 , 109)	**-52.1 (-65.2 , -39.0)**	**<0.01**
>499	15 (3 , 27)	**-140.0 (-154.4 , -125.6)**	**<0.01**
Previous AIDS diagnosis			
No	123 (120 , 126)	-	-
Yes	118 (113 , 124)	-5.6 (-13.5 , 2.3)	0.17
Site of care			
Outside London	115 (111 , 119)	**-9.6 (-15.9 , -3.3)**	**<0.01**
London	124 (120 , 127)	-	-
Age group			
15-24	129 (118 , 139)	**19.8 (5.3 , 34.2)**	**<0.01**
25-29	121 (114 , 128)	**11.5 (0.2 , 22.8)**	**0.05**
30-34	119 (113 , 125)	5.8 (-4.4 , 16.0)	0.26
35-39	119 (113 , 125)	3.6 (-6.6 , 13.7)	0.49
40-44	119 (112 , 126)	-	-
45-49	117 (108 , 127)	-4.6 (-17.7 , 8.6)	0.50
50-54	121 (108 , 135)	-2.0 (-18.5 , 14.5)	0.82
>54	124 (112 , 137)	-0.1 (-16.0 , 15.8)	0.99

Risk group			
MSM	139 (134 , 143)	-	-
Heterosexual men	101 (95 , 107)	**-26.2 (-37.1 , -15.4)**	**<0.01**
Heterosexual women	105 (100 , 110)	**-16.0 (-26.7 , -5.2)**	**<0.01**
Heterosexual women (diagnosed antenatally)	134 (124 , 145)	**35.4 (20.0 , 50.8)**	**<0.01**
IDU	90 (70 , 110)	**-44.5 (-67.3 , -21.8)**	**<0.01**
Recipients of blood products	88 (53 , 124)	-32.0 (-71.7 , 7.8)	0.12
Ethnicity			
White	135 (131 , 139)	-	-
Black African	103 (99 , 107)	**-33.1 (-43.3 , -22.8)**	**<0.01**
Black Caribbean	125 (110 , 140)	-5.9 (-24.4 , 12.6)	0.53
Black Other	111 (90 , 131)	**-26.8 (-50.9 , -2.8)**	**0.03**
Indian\Pakistani\Bangladeshi	103 (78 , 127)	-23.5 (-51.6 , 4.6)	0.10
Other Asian	118 (98 , 139)	**-25.0 (-48.8 , -1.1)**	**0.04**
Other/mixed	137 (122 , 152)	-0.9 (-18.4 , 16.5)	0.92
Overall	**120 (117 , 123)**	-	-
Comparison group*	-	**149.0 (132.3 , 165.7)**	-

* The comparison group was white MSM aged 40-44 years who had been diagnosed for more than 730 days and had a baseline CD4 cell count between 200 and 249 cells/mm^3 and no previous AIDS diagnosis when they started ART in 1999 in London.

Excluding individuals who started ART with CD4 cell counts greater than 349 cells/mm^3 reduced the number of individuals included in the model to 17,520. The effect of the year of starting ART was greater than in the main model such that initial rates of CD4 count increases in 2003 and 2004 were also significantly faster than in 1999 (p=0.04 and p<0.01 respectively) (Table 9.3). Individuals starting ART with a baseline CD4 count of 0-49 cells/mm^3 experienced a statistically slower rate of increase than those starting ART with CD4 counts of 200-249 cells/mm^3 (p=0.01). The effects of age group, heterosexual women

diagnosed antenatally and ‘Black other’ ethnicity were no longer associated with statistically different rates of increase than the comparison group.

Excluding women from the analysis reduced the number of individuals included in the model to 13,038 but had little effect on the results compared to the main model (Table 9.3). However, age group and ‘Black other’ ethnicity were no longer associated with different rates of increase in CD4 cell counts.

The analysis was repeated using different methods of calculating the baseline CD4 cell count (no change in the number of individuals included in the analysis). The baseline CD4 cell count categories changed for 14.4% (3,049) of individuals when only the last CD4 cell count was used and for 3.4% (721) of individuals when the mean of all CD4 cell counts in the last three months before ART was used. As the number of CD4 cell counts included in the calculation of the baseline value increased, there was an overall increase in the baseline CD4 count for individuals with CD4 counts below 250 cells/mm^3 and an overall decrease in the baseline CD4 count for individuals with CD4 counts of 250 cells/mm^3 or more such that there was a slight decrease in the mean and a reduced variance in baseline CD4 cell counts (from 226.2 cells/mm^3/34,160 when using only the last CD4 count, to 225.0 cells/mm^3/32,798 when using the mean of the last two CD4 counts where available, and to 224.7 cells/mm^3/32,469 when using the mean of all CD4 counts – all within the three months before starting ART).

Table 9.3. Results from linear mixed-effects analysis of the association between each factor and initial rates of increase of CD4 counts during the first three months of ART: sensitivity analyses of models i) excluding individuals starting ART with CD4≥350 cells/mm^3 and ii) excluding women.

	i) Excluding individuals starting ART with CD4≥350 cells/mm^3 N = 17,520		**ii) Excluding women N = 13,038**	
	Regression coefficient (95% CI)	**p value**	**Regression coefficient (95% CI)**	**p value**
Calendar year of starting ART				
1999	-	-	-	-
2000	1.6 (-15.8 , 19.0)	0.86	0.1 (-20.1 , 20.4)	0.99
2001	6.9 (-9.9 , 23.7)	0.42	7.5 (-12.4 , 27.4)	0.46
2002	13.7 (-2.4 , 29.9)	0.10	7.0 (-12.3 , 26.2)	0.48
2003	**16.0 (0.6 , 31.4)**	**0.04**	13.2 (-5.2 , 31.7)	0.16
2004	**23.6 (8.3 , 38.9)**	**<0.01**	17.2 (-1.1 , 35.5)	0.07
2005	**39.9 (24.8 , 55.1)**	**<0.01**	**39.5 (21.3 , 57.6)**	**<0.01**
2006	**47.0 (31.8 , 62.3)**	**<0.01**	**37.6 (19.5 , 55.7)**	**<0.01**
2007	**48.4 (32.8 , 64.0)**	**<0.01**	**47.3 (28.8 , 65.8)**	**<0.01**
Baseline CD4 count category (cells/mm^3)				
0-49	**-14.0 (-24.8 , -3.2)**	**0.01**	-5.8 (-20.9 , 9.3)	0.45
50-99	**14.3 (3.0 , 25.6)**	**0.01**	**21.4 (5.4 , 37.3)**	**<0.01**
100-149	7.4 (-3.5 , 18.4)	0.18	13.3 (-2.2 , 28.7)	0.09
150-199	3.5 (-6.6 , 13.7)	0.49	10.1 (-4.1 , 24.3)	0.17
200-249	-	-	-	-
250-299	**-17.5 (-28.4 , -6.5)**	**<0.01**	**-19.7 (-34.8 , -4.6)**	**0.01**
300-349	**-24.1 (-37.2 , -10.9)**	**<0.01**	**-26.9 (-45.8 , -8.0)**	**<0.01**
350-499			**-56.2 (-73.7 , -38.6)**	**<0.01**
>499			**-165.5 (-184.7 , -146.2)**	**<0.01**
Previous AIDS diagnosis				
No	-	-	-	-
Yes	-0.4 (-8.0 , 7.2)	0.92	-9.0 (-19.0 , 1.1)	0.08
Site of care				
Outside London	**-6.4 (-12.6 , -0.1)**	**0.05**	**-11.9 (-20.1, -3.7)**	**<0.01**
London	-	-	-	-

Age group				
15-24	10.0 (-5.0 , 25.0)	0.19	5.2 (-16.4 , 26.9)	0.64
25-29	5.6 (-5.7 , 16.9)	0.33	8.4 (-6.8 , 23.5)	0.28
30-34	7.3 (-2.7 , 17.3)	0.15	11.4 (-1.3 , 24.1)	0.08
35-39	1.9 (-8.0 , 11.8)	0.71	5.3 (-7.1 , 17.6)	0.40
40-44	-	-	-	-
45-49	-3.9 (-16.7 , 8.9)	0.55	-3.1 (-18.6 , 12.4)	0.69
50-54	-6.2 (-22.0 , 9.7)	0.45	-3.2 (-22.7 , 16.3)	0.75
>54	2.3 (-13.2 , 17.9)	0.77	4 (-14.4 , 22.4)	0.67
Risk group				
MSM	-	-	-	-
Heterosexual men	**-26.6 (-37.3 , -15.9)**	**<0.01**	**-31.4 (-44.4, -18.3)**	**<0.01**
Heterosexual women	**-19.4 (-30.1 , -8.6)**	**<0.01**		
Heterosexual women (diagnosed antenatally)	8.9 (-8.0 , 25.7)	0.30		
IDU	**-39.7 (-62.4 , -17.1)**	**<0.01**	**-43.9 (-67.3 , -20.5)**	**<0.01**
Recipients of blood products	-34.7 (-74.3 , 4.9)	0.09	-34.0 (-75.2 , 7.1)	0.11
Ethnicity				
White	-	-	-	-
Black African	**-32.5 (-42.8 , -22.3)**	**<0.01**	**-29.3 (-43.9 , -14.7)**	**<0.01**
Black Caribbean	-8.7 (-27.2 , 9.7)	0.35	-6.0 (-30.4 , 18.3)	0.63
Black Other	-15.2 (-38.9 , 8.4)	0.21	-14.2 (-49.1 , 20.8)	0.43
Indian\Pakistani\Bangladeshi	-8.1 (-35.5 , 19.3)	0.56	-19.6 (-54.0 , 14.8)	0.27
Other Asian	**-26.1 (-49.6 , -2.6)**	**0.03**	**-39.9 (-71.7 , -8.0)**	**0.01**
Other/mixed	0.2 (-17.1 , 17.6)	0.98	-5.8 (-26.7 , 15.0)	0.58
Comparison group*	**145.6 (129.0 , 162.3)**	-	**145.2 (125.1 , 165.2)**	-

* The comparison group was white MSM aged 40-44 years who had been diagnosed for more than 730 days and had a baseline CD4 cell count between 200 and 249 cells/mm^3 and no previous AIDS diagnosis when they started ART in 1999 in London.

In multivariable analysis, there was no meaningful difference between the models that used the mean of the last two (where available) CD4 cell counts or all CD4 cell counts within the previous three months (Table 9.4). Results from the model that used just the last CD4 count before ART indicated that the initial rates of increase in CD4 cell counts from baselines of 100-149 cells/mm^3 were

statistically faster than the increases from baselines of 200-249 cells/mm^3 in contrast to the main model (Table 9.4). There was little change in the initial rate of increase for any of the other variables.

The analysis was repeated using data for only the 15,509 individuals who had startART reported. These individuals had lower baseline CD4 cell counts (median 182 [IQR 90, 266] cells/mm^3) than the 5,694 individuals with estART but no startART reported (median 247 [IQR 136, 389] cells/mm^3) ($p<0.01$). For individuals with startART, the rate of change during the first three months of ART was 138 (95% CI 135, 141) cells/mm^3 compared to 58 (95% CI 52, 65) cells/mm^3 for those with estART. In comparison with the main model, there was no longer a significant association between age group and initial rates of CD4 cell count increases although the regression coefficients changed little. In contrast, the effect of site of care was no longer significant and the regression coefficient became close to zero. (Table 9.4). There was no change in the associations between risk group or ethnicity and initial rates of CD4 count increases. There was a weaker association with the year of starting ART and the baseline CD4 cell count category. Additionally, individuals with baseline CD4 counts less than 50 cells/mm^3 had significantly slower initial rates of increase, and individuals with baseline CD4 counts of 50-99 cells/mm^3 no longer had statistically different initial rates of increase, than the comparison group. However, a previous AIDS diagnosis was weakly associated with faster initial rates of CD4 cell count increase in this model ($p=0.05$).

Table 9.4. Results from linear mixed-effects analysis of the association between each factor and the rate of increase of CD4 counts during the first three months of ART: sensitivity analyses of models i & ii) varying the definition of baseline CD4 cell count and iii) only including individuals with startART reported.

	i) Single baseline CD4 cell count N = 21,203		**ii) Mean of all CD4 cell counts during the last three months N = 21,203**		**iii) Individuals with startART only N = 15,509**	
	Regression coefficient (95% CI)	**p value**	**Regression coefficient (95% CI)**	**p value**	**Regression coefficient (95% CI)**	**p value**
Calendar year of starting ART						
1999	-	-	-	-	-	-
2000	-1.8 (-19.1 , 15.4)	0.83	-1.0 (-17.8 , 15.7)	0.90	2.1 (-19.0 , 23.1)	0.85
2001	2.8 (-13.9 , 19.6)	0.74	2.7 (-13.5 , 19.0)	0.74	-4.2 (-24.6 , 16.3)	0.69
2002	2.8 (-13.4 , 18.9)	0.74	2.7 (-13.0 , 18.4)	0.74	-3.7 (-23.3 , 15.9)	0.71
2003	9.4 (-5.9 , 24.7)	0.23	8.9 (-6.0 , 23.7)	0.24	1.8 (-16.9 , 20.5)	0.85
2004	9.6 (-5.7 , 24.9)	0.22	11.0 (-3.9 , 25.8)	0.15	8.5 (-10.0 , 27.0)	0.37
2005	**28.9 (13.7 , 44)**	**<0.01**	**29.3 (14.6 , 44.0)**	**<0.01**	14.9 (-3.2 , 33.1)	0.11
2006	**35.3 (20.1 , 50.5)**	**<0.01**	**35.8 (21.1 , 50.6)**	**<0.01**	**23.7 (5.5 , 41.9)**	**0.01**
2007	**40.2 (24.6 , 55.8)**	**<0.01**	**39.5 (24.3 , 54.6)**	**<0.01**	**20.7 (2.3 , 39.0)**	**0.03**
Baseline CD4 count category (cells/mm^3)						
0-49	-6.5 (-18.5 , 5.5)	0.29	-10.8 (-22.4 , 0.8)	0.07	**-25.8 (-38.5 , -13.1)**	**<0.01**
50-99	**15.9 (3.3 , 28.4)**	**0.01**	**14.4 (2.2 , 26.7)**	**0.02**	9.4 (-4.0 , 22.8)	0.17
100-149	**15.8 (3.7 , 28.0)**	**0.01**	6.4 (-5.5 , 18.3)	0.29	1.5 (-11.6 , 14.5)	0.83
150-199	4.3 (-7.2 , 15.8)	0.47	3.9 (-7.1 , 14.9)	0.49	1.9 (-10.2 , 14.0)	0.76
200-249	-	-	-	-		
250-299	**-20.2 (-32.7 , -7.7)**	**<0.01**	**-19.6 (-31.4 , -7.8)**	**<0.01**	-13.1 (-26.3 , 0.1)	0.05
300-349	**-33.4 (-48.1 , -18.8)**	**<0.01**	**-29.1 (-43.3 , -14.9)**	**<0.01**	-7.3 (-23.9 , 9.3)	0.39
350-499	**-54.2 (-67.5 , -40.9)**	**<0.01**	**-48.0 (-61.0 , -35.0)**	**<0.01**	**-41.3 (-56.8 , -25.7)**	**<0.01**
>499	**-159.0 (-173.6 , -144.4)**	**<0.01**	**-138.2 (-152.5 , -123.8)**	**<0.01**	**-118.3 (-136.0 , -100.5)**	**<0.01**
Previous AIDS diagnosis						
No	-	-	-	-	-	-
Yes	-7.4 (-15.5, 0.7)	0.07	-5.2 (-13.1 , 2.7)	0.20	**9.0 (0.0 , 18.1)**	**0.05**
Site of care						
Outside London	**-9.2 (-15.6, -2.7)**	**<0.01**	**-9.7 (-16.0 , -3.5)**	**<0.01**	-0.5 (-7.6 , 6.6)	0.89
London	-	-	-	-	-	-

Age group						
15-24	**22.8 (8.0 , 37.6)**	**<0.01**	**19.3 (4.9 , 33.7)**	**<0.01**	13.3 (-3.1 , 29.6)	0.11
25-29	**13.4 (1.8 , 25)**	**0.02**	11.2 (-0.1 , 22.4)	0.05	7.0 (-5.6 , 19.6)	0.27
30-34	6.5 (-3.9 , 16.9)	0.22	5.5 (-4.6 , 15.6)	0.29	5.3 (-6.0 , 16.5)	0.36
35-39	3.7 (-6.7 , 14.1)	0.48	3.6 (-6.5 , 13.7)	0.49	1.5 (-9.7 , 12.7)	0.80
40-44	-	-	-	-	-	-
45-49	-4.1 (-17.5 , 9.4)	0.55	-4.6 (-17.6 , 8.5)	0.49	-8.5 (-23.1 , 6.1)	0.25
50-54	-1.1 (-18 , 15.8)	0.90	-1.5 (-17.9 , 15.0)	0.86	-6.8 (-25.1 , 11.6)	0.47
>54	-0.5 (-16.8 , 15.8)	0.95	0.1 (-15.7 , 15.9)	0.99	-5.3 (-23.0 , 12.3)	0.56
Risk group						
MSM	**-**	**-**	-	-	-	-
Heterosexual men	**-27.3 (-38.4 , -16.2)**	**<0.01**	**-26.0 (-36.8 , -15.2)**	**<0.01**	**-23.7 (-35.8 , -11.6)**	**<0.01**
Heterosexual women	**-16.8 (-27.8 , -5.9)**	**<0.01**	**-15.6 (-26.3 , -4.9)**	**<0.01**	**-17.0 (-28.9 , -5.2)**	**<0.01**
Heterosexual women (diagnosed antenatally)	**38.5 (22.8 , 54.2)**	**<0.01**	**34.2 (18.9 , 49.5)**	**<0.01**	**23.3 (6.0 , 40.6)**	**<0.01**
IDU	**-45.6 (-68.8 , -22.3)**	**<0.01**	**-44.6 (-67.2 , -22.0)**	**<0.01**	**-33.7 (-60.6 , -6.8)**	**0.01**
Recipients of blood products	-31.8 (-72.5 , 8.8)	0.13	-32.6 (-72.1 , 7.0)	0.11	-8.6 (-57.4 , 40.3)	0.73
Ethnicity						
White	-	-	-	-	-	-
Black African	**-33.9 (-44.4 , -23.5)**	**<0.01**	**-32.9 (-43.0 , -22.7)**	**<0.01**	**-33.2 (-44.6 , -21.9)**	**<0.01**
Black Caribbean	-6.3 (-25.2 , 12.6)	0.52	-6.4 (-24.8 , 11.9)	0.49	-17.1 (-37.3 , 3.1)	0.10
Black Other	**-26.1 (-50.7 , -1.6)**	**0.04**	**-26.4 (-50.3 , -2.5)**	**0.03**	**-29.7 (-56.8 , -2.5)**	**0.03**
Indian\Pakistani\Bangladeshi	-23.9 (-52.7 , 4.9)	0.10	-22.9 (-50.8 , 5.1)	0.11	-4.2 (-35.8 , 27.3)	0.79
Other Asian	-23.8 (-48.3 , 0.6)	0.06	**-25.3 (-49.0 , -1.6)**	**0.04**	**-29.8 (-55.8 , -3.8)**	**0.03**
Other/mixed	-1.5 (-19.4 , 16.4)	0.87	-0.7 (-18.0 , 16.6)	0.94	0.4 (-18.5 , 19.3)	0.97
Comparison group*	**151.8 (134.7 , 168.8)**	-	**149.0 (132.5 , 165.6)**	-	**166.9 (146.8 , 187.0)**	-

* The comparison group was white MSM aged 40-44 years who had been diagnosed for more than 730 days and had a baseline CD4 cell count between 200 and 249 cells/mm^3 and no previous AIDS diagnosis when they started ART in 1999 in London.

9.4.2 Rates of change of CD4 counts after the first three months of ART

Of the 21,203 individuals with baseline CD4 cell counts who started ART between 1999 and 2007, there were 18,657 (88.0%) who had CD4 cell counts after the first three months of ART (median 10 [IQR 5, 17; range 1, 71]). As in the previous analysis, these individuals were not representative of the 25,727 individuals starting ART (Table 9.5). In contrast to the previous analysis, however, individuals starting ART in 2007 were significantly less likely to be included in this analysis due to insufficient CD4 counts after the first three months of ART. Indian\Pakistani\Bangladeshi individuals were no longer significantly less likely to be included in the analysis whereas black Africans were significantly less likely to be included. Other differences in the factors associated with inclusion between the analyses were minor.

The overall mean rate of increase in CD4 counts after the first three months of ART was 43 (95% CI 42, 44) cells/mm^3 per year. This rate increased with year of starting ART but slowed with increasing baseline CD4 counts (Table 9.6). Individuals aged 15-24 years ($p<0.01$) and those aged 55 years or older ($p<0.01$) had slower rates of increase than those aged 40-44 years. Care in London ($p<0.01$) and a previous AIDS diagnosis ($p<0.01$) were associated with faster rates of increase in CD4 cell counts. Heterosexual men, heterosexual women diagnosed antenatally and IDU all had slower rates of increase than MSM whereas other heterosexual women had faster rates of increase than MSM. The only significant association with ethnicity was that black Africans had slower rates of increase in CD4 cell counts than white individuals.

Table 9.5. Individuals included in analysis of rates of increase in CD4 cell counts after the first three months after starting ART as a proportion of all individuals who started ART between 1999 and 2007.

	Number of individuals included in analysis	Proportion of all individuals starting ART (%)	Multivariable odds ratio for inclusion in analysis (95% confidence interval)
Calendar year of starting ART			
1999	1,288	69.8	-
2000	1,416	72.3	1.15 (1.00 , 1.32)
2001	1,650	74.2	**1.31 (1.14 , 1.50)**
2002	1,925	69.7	1.07 (0.94 , 1.21)
2003	2,676	78.9	**1.77 (1.55 , 2.02)**
2004	2,701	79.5	**1.85 (1.62 , 2.11)**
2005	2,848	82.7	**2.29 (2.00 , 2.62)**
2006	2,692	78.5	**1.72 (1.51 , 1.96)**
2007	1,461	44.6	**0.37 (0.33 , 0.42)**
Previous AIDS diagnosis			
No	14,300	72.1	-
Yes	4,357	74.0	**1.11 (1.03 , 1.19)**
Site of care			
Outside London	7,650	71.0	**0.85 (0.80 , 0.90)**
London	11,007	73.6	-
Age group			
15-24	1,288	69.5	**0.83 (0.72 , 0.94)**
25-29	2,888	71.2	**0.90 (0.81 , 1.00)**
30-34	4,415	73.8	1.03 (0.94 , 1.14)
35-39	4,169	74.4	1.06 (0.96 , 1.17)
40-44	2,719	72.9	-
45-49	1,509	71.8	0.95 (0.84 , 1.08)
50-54	793	70.6	0.87 (0.75 , 1.01)
>54	876	68.6	**0.77 (0.66 , 0.88)**
Risk group			
MSM	7,133	74.9	-
Heterosexual men	3,861	70.6	**0.80 (0.72 , 0.88)**
Heterosexual women	5,797	71.6	**0.85 (0.77 , 0.94)**
Heterosexual women (diagnosed antenatally)	1,362	72.6	0.89 (0.77 , 1.02)
IDU	379	64.7	**0.55 (0.46 , 0.66)**
Recipients of blood products	125	70.6	0.83 (0.59 , 1.16)

Ethnicity			
White	8,264	73.7	-
Black African	8,274	71.6	**0.89 (0.81 , 0.97)**
Black Caribbean	635	72.1	0.94 (0.80 , 1.12)
Black Other	339	69.3	0.85 (0.68 , 1.05)
Indian\Pakistani\Bangladeshi	238	69.0	0.83 (0.65 , 1.06)
Other Asian	302	75.1	1.11 (0.87 , 1.42)
Other/mixed	605	72.9	0.93 (0.79 , 1.10)
Total	18,657	72.5	-

Table 9.6. Results from linear mixed-effects analysis of the association between each factor and the rate of increase of CD4 counts after the first three months of ART (cells/mm^3 per year).

	Rate of increase of CD4 count estimated from univariable regression (95% CI) /cells/mm^3 per year	Multivariable regression coefficient (95% CI)	Multivariable p value
Calendar year of starting ART			
1999	23 (21, 26)	-	-
2000	28 (25, 31)	3.1 (-0.4 , 6.6)	0.08
2001	35 (32, 37)	**9.9 (6.5 , 13.3)**	**<0.01**
2002	40 (37, 42)	**14.4 (11.0 , 17.7)**	**<0.01**
2003	46 (44, 48)	**21.5 (18.3 , 24.8)**	**<0.01**
2004	56 (54, 59)	**31.1 (27.7 , 34.5)**	**<0.01**
2005	66 (63, 69)	**39.3 (35.4 , 43.2)**	**<0.01**
2006	79 (74, 85)	**52.4 (46.1 , 58.7)**	**<0.01**
2007	90 (61, 119)	**63.6 (35.6 , 91.7)**	**<0.01**
Baseline CD4 count category (cells/mm^3)			
0-49	62 (59, 64)	**14.2 (10.8 , 17.6)**	**<0.01**
50-99	55 (52, 58)	**9.5 (6.0 , 13.1)**	**<0.01**
100-149	50 (47, 52)	**5.4 (2.0 , 8.8)**	**<0.01**
150-199	50 (48, 53)	**3.6 (0.3 , 6.9)**	**0.03**
200-249	46 (44, 49)	-	-
250-299	42 (40, 45)	-1.8 (-5.4 , 1.7)	0.31
300-349	31 (27, 35)	**-10.6 (-14.7 , -6.4)**	**<0.01**
350-499	17 (14, 20)	**-23.3 (-27.0 , -19.5)**	**<0.01**
>499	-11 (-14, -8)	**-49.2 (-53.2 , -45.3)**	**<0.01**
Previous AIDS diagnosis			
No	39 (38, 40)	-	-
Yes	53 (51, 55)	**5.8 (3.6 , 8.0)**	**<0.01**
Site of care			
Outside London	45 (44, 46)	**2.2 (0.8 , 3.6)**	**<0.01**
London	41 (40, 42)	-	-

Age group			
15-24	28 (24, 31)	**-6.9 (-11.1 , -2.6)**	**<0.01**
25-29	36 (34, 38)	-2.8 (-6.0 , 0.5)	0.10
30-34	44 (42, 46)	1.7 (-1.2 , 4.6)	0.26
35-39	46 (44, 48)	1.0 (-2.0 , 3.9)	0.52
40-44	47 (45, 50)	-	-
45-49	49 (45, 52)	1.3 (-2.5 , 5.2)	0.50
50-54	43 (39, 48)	-3.6 (-8.4 , 1.1)	0.14
>54	39 (35, 44)	**-9.1 (-13.7 , -4.4)**	**<0.01**
Risk group			
MSM	42 (41, 44)	-	-
Heterosexual men	45 (43, 47)	**-4.6 (-7.7 , -1.6)**	**<0.01**
Heterosexual women	50 (49, 52)	**4.6 (1.6 , 7.6)**	**<0.01**
Heterosexual women (diagnosed antenatally)	8 (4, 12)	**-24.4 (-28.9 , -19.9)**	**<0.01**
IDU	26 (19, 32)	**-18.3 (-24.6 , -12.1)**	**<0.01**
Recipients of blood products	35 (24, 46)	1.7 (-8.8 , 12.1)	0.76
Ethnicity			
White	41 (40, 42)	-	-
Black African	44 (43, 45)	**-3.9 (-6.8 , -1.0)**	**<0.01**
Black Caribbean	49 (43, 54)	4.3 (-0.9 , 9.4)	0.10
Black Other	49 (42, 56)	0.8 (-6.1 , 7.7)	0.83
Indian\Pakistani\Bangladeshi	43 (35, 52)	0.5 (-7.5 , 8.6)	0.90
Other Asian	46 (39, 54)	-2.9 (-10.2 , 4.4)	0.44
Other/mixed	43 (38, 48)	-1.0 (-6.1 , 4.1)	0.70
Total	43 (42, 44)	**-**	**-**
Comparison group*	**-**	**28.1 (24.0, 32.3)**	**-**

* The comparison group was white MSM aged 40-44 years who had been diagnosed for more than 730 days and had a baseline CD4 cell count between 200 and 249 cells/mm^3 and no previous AIDS diagnosis when they started ART in 1999 in London.

Exclusion of those starting ART with CD4 cell counts greater than 350 cells/mm^3 (leaving 15,487 individuals in the dataset) had little effect on the results except that there was a more marked effect of year of starting ART and individuals aged 15-24 years no longer had significantly slower rates of increase (p=0.17). Exclusion of women (leaving 11,498 men in the dataset) also resulted in a more marked effect of the year of starting ART but reduced some of the effect of baseline CD4 cell counts. Men aged 15-24 years no longer had significantly different rates of increase than men aged 40-44 years (p=0.75) but men aged 50-54 years had significantly slower rates of increase (p=0.02). There was no longer any significant difference between the rates of increase for heterosexual men and MSM in the model that excluded women (p=0.19).

There was little change in the effect of any of the factors on the rate of increase when the analysis was repeated using either the last CD4 count before starting ART or the mean of all the CD4 counts in the last three months before starting ART as the baseline CD4 count.

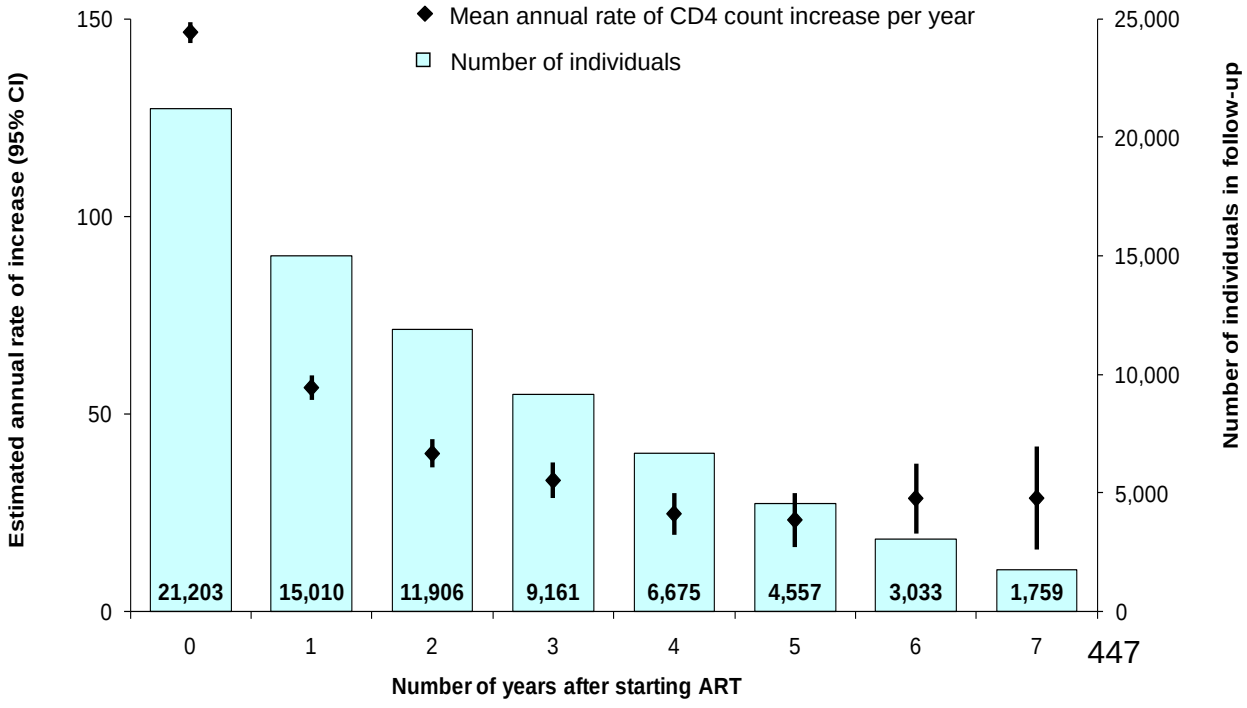
The analysis was repeated using only the records from individuals with startART. In this analysis of 14,045 individuals, the rate of change of CD4 counts after the first three months of ART was 47.2 (95% CI 46.1, 48.2) cells/mm^3 per year (compared to 29.7 (95% CI 27.8, 31.6) cells/mm^3 per year for individuals without startART reported). There was little difference between this model and the model including all individuals. Yet, there was no longer any significant difference between the rate of increase for heterosexual men and MSM (p=0.07) or between individuals treated at different sites of care (p=0.48).

9.4.3 How long after starting ART do baseline factors impact on CD4 cell count increases?

The 21,203 individuals with baseline CD4 cell counts who started ART post-1998 were included in this analysis. The number of individuals contributing to each separate analysis decreased from 21,203 for the first year of follow-up to 1,759 for the seventh year of follow-up (Figure 9.3).

The mean rate of increase declined progressively over the first five years after starting ART but subsequently remained fairly stable and positive (Figure 9.3). The mean increase in CD4 cell counts was: 147 (95% CI 144, 149) cells/mm^3, 57 (95% CI 54, 60) cells/mm^3, 40 (95% CI 36, 43) cells/mm^3, 33 (95% CI 29, 38) cells/mm^3, 25 (95% CI 19, 30) cells/mm^3, 23 (95% CI 16, 30) cells/mm^3, 29 (95% CI 20, 37) cells/mm^3, and 29 (95% CI 16, 42) cells/mm^3 in years 0, 1, 2, 3, 4, 5, 6 and 7, respectively.

Figure 9.3. Estimated mean (95% CI) rate of CD4 cell count increase per year and number of individuals in follow-up by year after starting ART.



Separate multivariable analysis for each year of follow-up indicated that, for the reference group who started ART in 1999 with CD4 counts of 200-249 cells/mm^3, the rate of increase declined from 178 (95% CI 163, 192) cells/mm^3 in the first year to 29 (95% CI 8, 50) cells/mm^3 in the fourth year, after which only the increase in the seventh year was significant (Table 9.7).

Many of the significant associations in multivariable analysis were observed during the first year after starting ART, which may reflect effects during the first three months. Individuals with baseline CD4 counts of 300 cells/mm^3 or more experienced slower rates of increase during the first three years after starting ART. This association appeared to remain longer for individuals with baseline CD4 counts of 500 cells/mm^3 or more. IDU tended to have slower rates of increase than MSM although the regression coefficients fluctuated between positive and negative and were only significant (negative) in the first, second and fifth years after starting ART. Heterosexual women diagnosed antenatally and recipients of blood products did not have faster rates of increase than MSM in any year after starting ART. Other heterosexuals had slower rates of increase in the first year but this was not maintained. Initiation of ART after 1999 was generally associated with faster rates of increase in CD4 cell counts although, during the first year after starting ART, rates appeared to be slower for individuals starting ART between 2000 and 2003 and faster after 2003. Black Africans appeared to experience slower rates of increase than white individuals although the difference was only significant during the first and fifth year after starting ART. In the first year after starting ART, there was some evidence that individuals aged 15-39 years had faster rates of increase in CD4 counts than

individuals aged 40-44 years and that individuals aged 45 years or older had slower rates of increase but there were no consistent associations in other years. There were other significant rates of change in CD4 counts in varied groups and varied years but no other consistent trends.

Table 9.7. Results from multivariable linear regression analysis of the association between each factor and the rate of increase of CD4 counts in each year of follow-up after starting ART.

	Regression coefficient (cells/mm^3), p value							
Number of years after starting ART	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Calendar year of starting ART								
1999	-	-	-	-	-	-	-	
2000	**-16, 0.03**	7, 0.33	1, 0.92	0, 0.99	16, 0.06	11, 0.21	3, 0.80	-24, 0.20
2001	-12, 0.09	**16, 0.02**	12, 0.10	5, 0.56	14, 0.11	**23, <0.01**	-14, 0.42	
2002	**-15, 0.03**	1, 0.87	2, 0.75	**16, 0.05**	**17, 0.04**	25, 0.10		
2003	-11, 0.10	**16, 0.02**	3, 0.68	**15, 0.05**	14, 0.28			
2004	4, 0.56	12, 0.06	**16, 0.02**	7, 0.53				
2005	10, 0.12	**18, <0.01**	-10, 0.37					
2006	**27, <0.01**	9, 0.33						
2007	**81, <0.01**							
Baseline CD4 count category (cells/mm^3)								
0-49	10, 0.07	**15, 0.01**	3, 0.65	6, 0.49	5, 0.66	-4, 0.78	-10, 0.60	40, 0.16
50-99	-1, 0.86	5, 0.46	4, 0.63	1, 0.88	3, 0.80	8, 0.60	-15, 0.47	42, 0.18
100-149	-1, 0.86	-3, 0.58	-4, 0.58	5, 0.58	5, 0.66	-5, 0.72	-24, 0.20	34, 0.25
150-199	4, 0.48	-4, 0.53	-1, 0.88	-12, 0.18	4, 0.72	-3, 0.82	-21, 0.27	15, 0.62
200-249	-	-	-	-	-	-	-	-
250-299	-11, 0.06	-8, 0.18	-4, 0.60	4, 0.67	4, 0.73	-14, 0.34	-4, 0.82	3, 0.91
300-349	**-46, <0.01**	**-19, 0.01**	**-17, 0.05**	-7, 0.50	-21, 0.11	17, 0.30	-41, 0.06	63, 0.06
350-499	**-79, <0.01**	**-39, <0.01**	**-15, 0.05**	-13, 0.20	11, 0.38	13, 0.39	8, 0.67	42, 0.16
>499	**-154, <0.01**	**-98, <0.01**	**-43, <0.01**	**-21, 0.04**	-3, 0.85	**-30, 0.05**	**-48, 0.02**	22, 0.49
Previous AIDS								
No	-	-	-	-	-	-	-	-
Yes	5, 0.19	**9, 0.03**	-4, 0.34	-7, 0.19	6, 0.4	8, 0.32	6, 0.58	-4, 0.81
Site of care								
Outside London	-5, 0.07	-1, 0.85	**8, 0.04**	**12, 0.01**	-1, 0.81	-4, 0.62	-1, 0.89	-15, 0.32
London	-	-	-	-	-	-	-	-

Age group								
15-24	9, 0.17	-6, 0.40	-15, 0.12	-2, 0.83	20, 0.16	26, 0.16	**-70, <0.01**	-51, 0.16
25-29	**15, <0.01**	0, 0.98	**-18, 0.01**	-8, 0.35	6, 0.54	23, 0.09	-31, 0.08	-21, 0.42
30-34	9, 0.06	4, 0.40	-5, 0.42	-11, 0.14	16, 0.09	**33, <0.01**	-6, 0.70	-9, 0.68
35-39	**9, 0.05**	0, 0.96	2, 0.74	-9, 0.27	2, 0.80	**31, 0.01**	-29, 0.07	7, 0.77
40-44	-	-	-	-	-	-	-	-
45-49	-12, 0.06	9, 0.17	-6, 0.46	-3, 0.73	11, 0.39	**34, 0.03**	7, 0.73	-49, 0.10
50-54	-14, 0.06	**-19, 0.03**	-5, 0.60	-10, 0.4	21, 0.15	34, 0.07	-14, 0.55	15, 0.66
>54	-12, 0.10	-9, 0.29	**-23, 0.02**	-14, 0.23	-26, 0.08	14, 0.45	23, 0.36	7, 0.85
Risk group								
MSM	-	-	-	-	-	-	-	-
Heterosexual men	**-32, <0.01**	1, 0.92	7, 0.30	4, 0.58	0, 0.97	-4, 0.77	32, 0.04	**57, 0.01**
Heterosexual women	**-11, 0.02**	**13, 0.02**	10, 0.13	4, 0.61	10, 0.30	4, 0.73	**37, 0.02**	**85, <0.01**
Heterosexual women (diagnosed antenatally)	10, 0.18	-14, 0.09	-10, 0.32	6, 0.64	-12, 0.45	-18, 0.39	-4, 0.90	86, 0.17
IDU	**-52, <0.01**	**-26, 0.03**	-3, 0.80	20, 0.26	**-56, <0.01**	18, 0.45	-21, 0.52	-17, 0.70
Recipients of blood products	-26, 0.14	21, 0.30	-14, 0.52	-31, 0.24	18, 0.57	-6, 0.88	69, 0.08	74, 0.24
Ethnicity								
White	-	-	-	-	-	-	-	-
Black African	**-19, <0.01**	-2, 0.70	-3, 0.63	0, 1.00	**-18, 0.05**	-14, 0.24	-10, 0.53	-41, 0.07
Black Caribbean	-4, 0.60	11, 0.24	15, 0.18	**38, <0.01**	12, 0.48	-33, 0.12	1, 0.98	-32, 0.47
Black Other	-14, 0.20	**32, 0.01**	-3, 0.81	14, 0.42	**-73, <0.01**	-6, 0.84	-37, 0.33	**-118, 0.02**
Indian\Pakistani\ Bangladeshi	-8, 0.52	-15, 0.28	**33, 0.05**	-20, 0.36	-40, 0.11	**-74, 0.02**	-58, 0.17	-12, 0.83
Other Asian	-13, 0.24	-12, 0.38	6, 0.69	5, 0.79	**-52, 0.02**	47, 0.14	34, 0.36	-25, 0.67
Other/mixed	-3, 0.70	-4, 0.64	12, 0.29	16, 0.22	-20, 0.23	-36, 0.06	3, 0.92	-35, 0.36
Overall	**178, <0.01**	**50, <0.01**	**40, <0.01**	**29, <0.01**	10, 0.43	-6, 0.69	**48, 0.01**	2, 0.93

Significant differences in the first year after starting ART were similar to those during the first 91 days (Section 9.4.1). There were greater rates of increase in later years, among younger individuals and among MSM. In addition, there were slower rates of increase among individuals with higher baseline CD4 counts, among IDU and heterosexuals (not women diagnosed antenatally), and among individuals of black African ethnicity in both models. However, in contrast, the region of care was not significantly associated with rates of increase in CD4 cell counts after starting ART in this model.

9.5 Discussion

These analyses describe the population-level effect of ART on CD4 cell counts in the UK as a mean rate of increase of 120 cells/mm^3 during the first three months of ART and a subsequent rate of increase of 43 cells/mm^3 per year. Multivariable analyses indicated that there were statistically significant differences in the rate of increase between groups during the first three months of ART but sensitivity analysis using only data from individuals with startART reported suggested that many of these effects could be due to bias in the estimations of baseline CD4 cell counts at estART. In that analysis, baseline CD4 cell count, risk group, ethnicity and no previous AIDS diagnosis were significantly associated with CD4 cell count increases. Furthermore, it should be noted that combinations of factors could actually result in a decrease in CD4 cell counts during the first three months after starting ART in all models.

Analysis of the data stratified by year of follow-up indicated that the overall mean rate of increase remained greater than zero up to eight years after starting ART but declined with more years of follow-up. However, there was an inherent bias towards faster CD4 count increases because some of those who died or were lost to follow-up would be a result of poor response to therapy or disengagement with care. Furthermore, the representativeness of the data decreases with length of follow-up because only individuals starting ART more than five years ago, for example, can inform the five-year trend and these individuals will probably receive a different treatment history and different clinical care than individuals starting in 2007.

Rates of increase in CD4 counts of 120 cells/mm^3 during the first three months after ART were of the same order but slightly higher than those reported in the literature. This was despite the systematic bias from estimation of dates of starting ART, which was likely to have excluded some of the period when CD4 cell counts were rising most rapidly for many individuals. A systematic overview of 53 clinical trials determined a weighted mean increase in CD4 counts after 24 weeks of ART of 130 cells/mm^3 [515]. Cohort studies that followed ART-naïve individuals reported median increases in CD4 counts of 88 cells/mm^3 during the first eight weeks[513]; 80 cells/mm^3 during the first three months[529]; 114 cells/mm^3 during the first six months[498]; and 119 cells/mm^3 during the first six months[512]. As the best estimate of the inflection point after the period of rapid increase is ten weeks (among ART-experienced, but HAART-naïve, individuals) it is likely that about half of the follow-up time in the studies which determined increases over the first six months of ART was not during the period of rapid increase[523].

The rate of increase in CD4 counts after the first three months of ART in this analysis was 43 cells/mm^3 per year. This compares to an increase of 94 cells/mm^3 per year in a cohort study[498]. However, that was observed during a median follow-up of 2.5 years and studies which split the follow-up time indicate that the rate of increase slows over time[364;516;529]. For example, cohort data show that increases slowed from 172 cells/mm^3 during the first year after ART to 80, 63, 28, and 27 cells/mm^3 during the second to fifth years after ART[529]. Similarly, in the integrated dataset, the increase in CD4 counts slowed from 147 cells/mm^3 during the first year after ART to 57 cells/mm^3 in the second year, and then more gradually to around 25 cells/mm^3 in years 5-8.

Comparison of the results of these analyses with the literature is complicated because some publications focus on the effect of HAART and include, but do not adjust for, a high proportion of individuals who were pre-treated with non-HAART regimens[533;534]. Additionally, some studies focused on individuals who maintained low viral loads on ART rather than the entire population[364;514;516]. Studies also varied in the factors considered, the number of individuals monitored and the time periods after ART that were included in analyses, which may explain some of the differences in the associations found.

Consistent with these analyses, older age has previously been found to be associated with slower rates of increase of CD4 cell counts over both the short[516;531] and long-term[513;516;532;533] (although not when baseline CD8 cell counts were considered[498]). However, my analysis only showed a clear association with age in the short-term. Age may be biologically associated with CD4 cell count response to ART but it may also reflect other characteristics of the population, such as adherence.

Only one other study has clearly shown a reduced immunological response for IDU[531] although this does not seem to be refuted by other studies that exclude pre-treated individuals and is consistent with other information about reduced adherence, irregular attendance and poorer clinical outcomes among IDU[124].

These analyses have identified significant differences between MSM and heterosexuals that have not been reported in the literature. Heterosexual women diagnosed antenatally appeared to have greater increases in CD4 cell

counts during the first three months, which may have been due natural increases after delivery for women who started ART during pregnancy[57;509;535]. However, these were not sustained and overall rates of increase after the first three months were even slower than among MSM. This may have been due to interruption of therapy among women who were indicated ART solely for the prevention of mother-to-child transmission, or who interrupted for other reasons after giving birth. Other heterosexuals had slower rates of increase in CD4 counts than MSM during the first three months and the first year after starting ART but this did not appear to continue in the medium term. Explanations for this require further investigation although differential adherence and differential virulence of subtypes may be implicated (sections 1.3.3 and 1.3.6 respectively).

Results from the literature differed in the effect of baseline CD4 cell counts on the immunological response during the short-term after ART. Some studies indicated no effect of baseline CD4 count[486;498;513], whereas others observed an association between lower baseline CD4 counts and increased rates of change[515;531;532]. Results have also been inconclusive during long-term ART[498;533]. However, different authors adjusted for many different factors, which may have substantially affected the results and made them less comparable[*].

Most studies demonstrated no effect of sex on the immunological response to ART[498;531;533]. Two studies did find an association but did not consider risk

[*] Including time-updated viral loads, drug classes included in the treatment regimen at baseline, baseline CD8 cell counts, and time from seroconversion.

group as a factor[513;516]. One of these also found no effect of ethnicity over the short-term[513] whereas another study found that rates of increase of CD4 cell counts were higher among white individuals than non-white individuals during the first 3 months of ART but not subsequently[498]. These findings are supported by my analyses. The results of the latter study indicated that the difference between ethnicities was not due to a lower baseline viral load among black African individuals, which may have been a biological hypothesis. Black Africans were the only ethnic group in the integrated dataset to have significantly weaker immunological response after the first three months of ART, but it is likely that this may be due to unmeasured confounders such as differential adherence or virulence of subtypes.

Results from studies have also differed as to whether rates of increase of CD4 cell counts have increased over time or not[515;531]. This may be because there is no effect of when individuals started ART on the immunological response during the first 3 months of ART but that this has had a significant effect after the first 3 months of ART[498]. Recent calendar years were associated with greater rates of increase over the short-term in my main analysis but not in the sensitivity analysis considering only records with startART. This suggests that the observed effect of calendar year over the short-term was due to bias in the estimation of estART. Except for 1999, there did not appear to be a strong effect of baseline calendar year over the long term.

An AIDS diagnosis prior to starting ART was associated with rates of increase in the sensitivity analysis excluding records with estART but not in the main

analysis of the first three months after ART and there were also conflicting reports in the literature[531;532]. Interestingly, a previous AIDS diagnosis was associated with greater increases in CD4 cell counts rather than poorer immunological responses, which may reflect greater preventative adherence among individuals who have experienced an HIV-related illness.

There have not been any published results describing whether immunological response to ART varies across the UK but there have been some suggestions (also refuted) that treatment in tertiary centres by specialist clinicians who care for large cohorts of HIV-infected individuals and who have greater and earlier experience with new drugs due to involvement in clinical trials may result in improved prognosis[536-540]. In these analyses, individuals who were treated in London had greater rates of CD4 cell count increase during the first three months after starting ART but there were no longer-term regional differences. However, the regional difference detected may be a result of bias due to less complete reporting of startART from outside London as there was no significant difference by region of treatment when excluding records with estART.

There were several factors that have been associated with the immunological response to ART but are not within the scope of these analyses. Most significantly, higher baseline viral loads are associated with greater CD4 cell count increases both in the short and long-term[498;513;516;531;533]. Although a systematic overview found a faster rate of CD4 increase among individuals treated with a boosted PI, the authors acknowledged that they may not have been able to fully differentiate the baseline characteristics of those initiating

different regimens, and most other studies have found no effect of the initial treatment regimen on CD4 response[481;498;513;515;531;532]. Lower pre-HAART CD8 cell counts, pre-treatment slopes and a higher baseline naïve/memory CD4 cell count ratio have also been associated with a greater increase in CD4 counts[485;498;513;533]. The effectiveness of treatment after initiation also affects CD4 count responses and discontinuation of treatment and virological failure result in significantly less benefit[486;541;542].

The analysis of rates of increase in CD4 cell counts during the first three months after ART was likely to have been affected by inclusion bias because the analysis included an unrepresentative 82.4% of all individuals who started ART during the period. It is not possible to know how individuals included in each population group differed from those excluded, according to this outcome of CD4 changes after ART initiation, so interpretation must be based on the assumption that they are similar in this regard. White individuals, MSM and Londoners were more likely to be included in the integrated dataset and more likely to have CD4 counts for analysis than other groups. These groups also had faster CD4 count increases than other groups indicating that the overall population results would be significantly biased in this direction. Older individuals were less likely to be included in the integrated dataset, were less likely to have CD4 counts for this analysis, and had slower CD4 count increases than other groups, which would also add to this bias and result in overestimation of the overall rate of increase of CD4 cell counts. Therefore, exclusion bias may partly account for the slightly higher rates of increase in CD4 cell counts during the first three months after starting ART.

Selection bias may have been slightly greater in the analysis after the first three months of ART because individuals without CD4 cell counts more than three months after starting ART were not included in that analysis (72.5% versus 82.4%). This particularly resulted in under-representation of individuals starting ART in 2007 who were more likely to have faster CD4 count increases. Black Africans were also particularly less likely to be included in this analysis and were already under-represented in the integrated dataset and shown to have slower CD4 cell count increases. MSM and Londoners were again over-represented in the dataset and shown to have faster CD4 count increases than other groups. The balance of these effects was likely to be that the overall results were biased towards a faster increase in CD4 counts.

The factors associated with inclusion in the analyses suggest that unmeasured factors also added to the bias. Individuals more likely to be included were more likely to have been monitored more frequently than other individuals (chapter 5), which may be due to their care-seeking behaviour, their risk of disease progression, their risk of clinical problems such as toxicity or the clinical practice of their doctor. All of these unmeasured factors are likely to be confounders due to their association with adherence and treatment response.

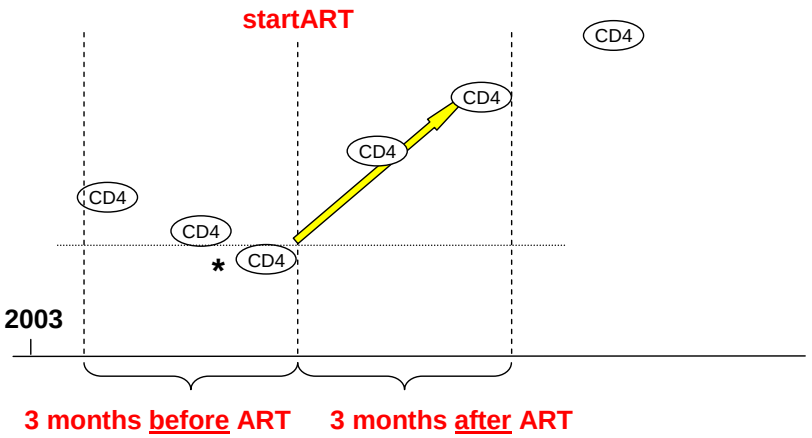
Informative drop-out (associated with the outcome) due to death may have occurred if those who died and were not included had poorer immunological responses to ART than other individuals. This may have resulted in over-estimation of the rates of CD4 cell count increases in all sub-groups in the

analysis and particularly for those with higher mortality rates, such as older individuals, IDU and individuals with low baseline CD4 cell counts.

The interruption of ART by women who had used it solely to reduce mother-to-child transmission may have biased the results, particularly in the longer-term, because their CD4 cell counts would subsequently be expected to fall. Exclusion of all women in sensitivity analysis (because pregnancy status was only known at the time of HIV diagnosis) reduced some of the effects such that the results were similar to the analysis excluding individuals with estART.

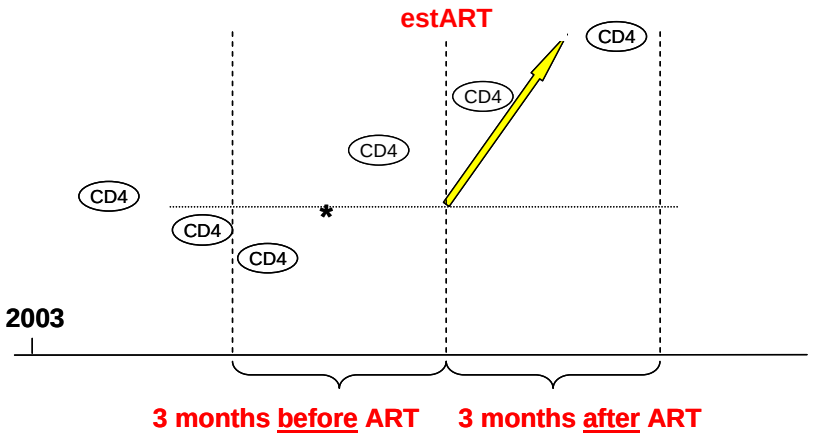
Error in the determination of the dates of starting ART (chapter 7) is likely to have had a substantial impact on the analysis during the first three months of starting ART. The effects of many factors were no longer statistically significant in the sensitivity analysis that excluded records without startART. Misclassification could have reduced or inflated the estimates of the rates of increase and may have biased the estimates of the regression coefficients for the factors investigated (Figures 9.4 a-c. An accurate schematic from a record with startART is shown in Figure 9.4a). Errors in the determination of estART were likely to have biased rates of CD4 increases because they would likely to be calculated using the wrong CD4 cell counts; for example, those after the initiation of ART (Figure 9.4b). Greater errors were likely if the baseline CD4 cell count could only be estimated from one CD4 count before estART rather than averaged between that CD4 count and a CD4 count closer to the true baseline (Figure 9.4c).

Figure 9.4a. Schematic showing the estimation of the slope during the first three months of ART for an individual with startART.



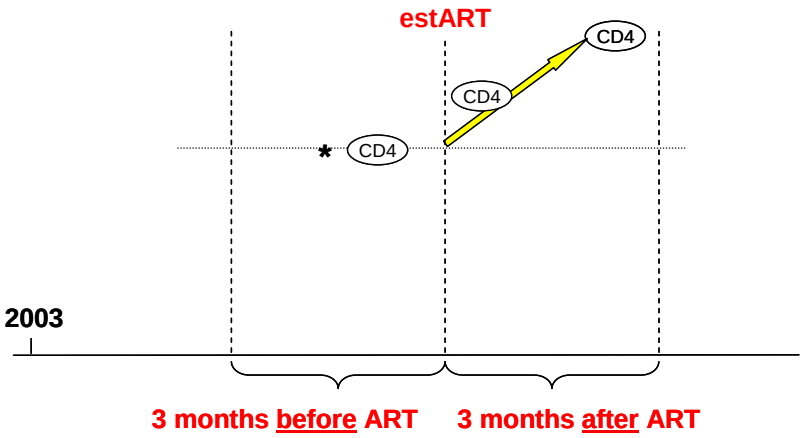
*** Baseline CD4 cell count was the average of the last two CD4 cell counts before startART/estART where available, otherwise the last CD4 cell count**

Figure 9.4b. Schematic showing the estimation of the slope during the first three months of ART for an individual with estART and more than one CD4 cell count before estART.



*** Baseline CD4 cell count was the average of the last two CD4 cell counts before startART/estART where available, otherwise the last CD4 cell count**

Figure 9.4c. Schematic showing the estimation of the slope during the first three months of ART for an individual with estART and only one CD4 cell count before estART.



*** Baseline CD4 cell count was the average of the last two CD4 cell counts before startART/estART where available, otherwise the last CD4 cell count**

A common source of error in studies of the immunological response to ART is regression to the mean, introduced by natural and measurement variability of the baseline CD4 cell counts[43;242;499;543-545]. For example, one study that included follow-up before and after treatment found that the decrease in CD4 cell counts prior to ART was significantly associated with subsequent increases to three and six months post-ART but not after that, but this may have been due to regression to the mean[529]. In these analyses, regression to the mean was likely to have affected the observed rate of increase during the first three months after starting ART and during the first year of ART because the baseline CD4 cell count was used in the calculation of the rate. In these analyses, the effect will have been greatest for individuals with baseline CD4 cell counts in the highest and lowest categories because the variability was more likely to have

acted in one direction (assuming that it was random and normally distributed) [546]. For example, individuals with observed baseline CD4 cell counts between 0 and 49 cells per mm^3 may have had true baseline CD4 cell counts greater than 49 cells per mm^3 but cannot have had true baseline CD4 cell counts below 0 cells per mm^3. The true overall rate of increase for this category of individuals was therefore expected to be slower than observed such that it was significantly slower than for individuals with baseline CD4 counts of 200-249 cells per mm^3. Use of the mean of the last two CD4 cell counts during the three months prior to starting ART was intended to reduce the effects of natural and measurement variability and regression to the mean[546]. As the number of CD4 cell counts included in the calculation of the baseline value increased, there was an increase in the baseline CD4 count for individuals with low CD4 counts and a decrease in the baseline CD4 count for individuals with high CD4 counts. This supported the assumption that there was an effect of regression to the mean. Because CD4 counts would be expected to be decreasing over time prior to starting ART, use of the mean of the last two measurements before starting ART may have systematically overestimated baseline values. In fact, there was a marginal increase in the overall medians when using the mean of the last two CD4 cell counts before ART (196 cells/mm^3) and the mean of all CD4 cell counts in the last three months before ART (197 cells/mm^3) compared to using just the last CD4 cell count before ART (195 cells/mm^3). However, this was likely to have less impact than regression to the mean and negligible impact on the relative effects of the baseline CD4 count categories.

Confounding by indication probably affected the results because individuals starting ART with CD4 cell counts below 200 cells/mm^3 should have been recommended initiation of ART according to treatment guidelines[134;426]. The unmeasured characteristics of these individuals were therefore likely to differ from those of individuals who started ART at higher CD4 cell counts.

In the stratified analysis of long-term follow-up, there may have been insufficient numbers of sub-groups of individuals to achieve statistical significance after the first three years (although overall numbers in the seventh year still exceeded 3,000). This was suggested by the fact that the slower rates of change for IDU and individuals aged 55 years or more continued throughout the eight years of follow-up but with decreasing significance.

Finally, it should be noted that more experienced statisticians may have used more complex techniques to analyse the data in a single random effects model. Inclusion of join points would have allowed for changes in the rate of increase in CD4 cell counts and changes in interactions between each baseline factor and the time from starting ART. A join point after 3 months on ART could have been used to merge the two formal models that were analysed, and multiple joins at each year after starting ART could have been used to formally test the effects of baseline covariates over time.

9.6 Conclusion

Advantages of analysing immunological responses using data from the integrated HIV surveillance systems were that the numbers of individuals

starting ART were substantially greater than numbers included in cohort analyses in the literature and that follow-up was prospective, extensive and continuing. This enabled multivariable analysis of factors such as ethnicity and risk group together with narrow CD4 cell count categories and calendar year of starting ART. Areas for improvement of surveillance data included more accurate dates of starting ART, pregnancy status as a reason for starting ART, dates of stopping or interrupting ART and viral loads at the time of starting ART.

These analyses identified the disadvantageous effect of black African, Indian/Pakistani/Bangladeshi and Other Asian ethnicity on the rate of increase of CD4 counts during the first three months after starting ART in E,W&NI. The consistency of this finding through sensitivity analyses and confirmation with the literature[498] indicate that it was unlikely to be due to the limitations of data. Wider research is required to investigate whether this is due to lower adherence during this time, which may be due to language barriers reducing understanding of clinical guidance, greater perception of stigma or competing priorities.

Other findings of this population-level analysis confirmed those in the literature. These include the overall short and long-term rates of change of CD4 cell counts, which increase rapidly during the first three months after starting ART and continue to increase at a slower rate for eight years (although biased by loss to follow-up, stopping treatment and death). The results of these analyses may help to clarify the population groups that may expect CD4 cell counts to recover to levels observed among uninfected individuals and to identify those that may require enhanced adherence or other support such as IDU.

The results showed that immunological responses were slower at higher CD4 counts and that the threshold at which rates of increase slowed varied according to sensitivity analyses. Maintaining higher CD4 counts reduces morbidity and mortality in the short-term but the optimal threshold at which to start ART is not yet known. Therefore, the randomised control trial, 'START'[127], is underway to determine whether deferral of ART can be clinically beneficial.

The results suggest an improved immunological response over time, and particularly in the long-term, but also highlight the need to ascertain accurate dates of starting ART for all individuals for increased reliability of results. This and the previous chapter demonstrate that this key piece of epidemiological data cannot be adequately determined by proxy from other surveillance data.

The methodology used in this analysis of surveillance data provides potential for the routine monitoring of the rate of increase of CD4 cell counts as a performance indicator of high quality clinical care. Increased awareness of the factors associated with slower rates of increase and direct feedback of these results could lead to changed protocols to address these issues. Furthermore, further analysis could quantify associations between slower rates of immunological increase and subsequent mortality.

In conclusion, understanding of the rates of increase of CD4 cell counts after starting ART and their determining factors in E,W&NI may be used by clinicians to benchmark patient-specific immunological responses to therapy. The public health benefits of this may be improved outcomes for HIV-infected patients.

Chapter 10. Conclusion

10.1 Summary

The rationale behind this thesis was to develop robust methodologies to realise some of the untapped public health potential of HIV surveillance data. Case-based HIV surveillance in E,W&NI is very comprehensive but analysis is limited because it consists of three complementary but independent systems due to the complexity of the epidemic and the separate primary sources of data. Therefore, the pivotal aim of this thesis was to develop a robust method to fully integrate data from these systems into a single national HIV surveillance database containing a complete and consistent description of each individual with a coherent sequence of events. The secondary aim was to apply statistical techniques, more commonly used in cohort study research, to the integrated database to demonstrate the full public health potential of the data.

Reliable analysis and astute interpretation of the data was dependent upon a thorough understanding of the HIV epidemic, its development over time, and the mechanisms through which the data are collected and processed. Chapter 1 describes the clinical and epidemiological background to HIV surveillance and chapter 2 details the three case-reporting systems and provides summary context of HIV surveillance systems in other countries with similar epidemics.

Chapters 3 and 4 describe the complex processes that were developed over a long period of time to achieve a single, national, integrated HIV surveillance database. After characterisation of patient-specific information in each system, I

created bilateral, deterministic record linkage algorithms that reliably matched patient records between HARS and SOPHID and between SOPHID and CD4 Surveillance. This demonstrated a degree of coding errors in the data and that a substantial number of records could not be automatically linked. This informed the selection of records for the triangulation process of Chapter 4 but raised issues of inclusion bias, data quality, disharmonised surveillance systems, under-reporting and loss to follow-up. I concluded that at least two of soundex, date of birth and local patient ID number should match for record linkage to be accepted without further manual review. Outcomes of these analyses were that adjustments for under-reporting are no longer used in E,W&NI to prevent over-counting, SOPHID uses integrated patient and report tables that facilitate analyses and data completeness, record linkage is used to reduce the follow-up needed for missing information, and record linkage is incorporated into data processing procedures to validate HIV case reports as they are received. Chapter 3 also demonstrated that there is a need for care when combining reports from parallel surveillance systems (including the development of novel systems as for the surveillance of recent HIV infections) because of the potential creation of additional duplicates. Results of this chapter can inform record linkage processes that are likely to be increasingly used in both surveillance and research because they permit new analyses without additional and duplicated data collection. The minimal additional resources required at national level could be offset by reductions in data collection and entry resources at all levels by reducing duplication.

Due to tripartite HIV case reporting systems, triangulation and rationalisation of common data variables was necessary to create the final integrated dataset as described in chapter 4. For consistency throughout this thesis, I chose to only use data that were linked across all three surveillance systems, which limited analyses to individuals alive in 1995 and reported to SOPHID and did not provide a complete historical picture. This would not be necessary for future analyses that do not require data points from all three systems. The resulting integrated dataset utilized and combined information from each surveillance system to reduce missing information and expand the dataset. Furthermore, the integrated dataset also identified additional data inconsistencies raising the question of whether redirection of resources could significantly improve data quality at minimal cost to data completeness. The final integrated dataset had a large sample size equivalent to multinational cohort collaborations, potential follow-up from HIV diagnosis to death, good information on ethnicity and national representativeness. Lessons learnt could inform the development of national cohorts for what could be termed '3rd generation HIV surveillance' as pioneered by the Netherlands and Austria.

A thorough understanding of any dataset is needed before analyses can be appropriately carried out and results can be interpreted. Some understanding of the integrated dataset was developed during the processes of record linkage, triangulation and rationalisation but these also raised issues of potential bias. Chapter 5 described the analyses of the integrated dataset undertaken to investigate potential sources of bias. Strengths of the integrated dataset were that inclusion bias could be specifically characterised as it was a subset of

national surveillance data, in which loss to follow-up may have been largely limited to death and emigration rather than transfer between clinics, and that follow-up started from the time of HIV diagnosis. A weakness was that record linkage incorporated additional errors into the data suggesting that misclassification bias may have affected these data more than clinical cohort data. Left and right censoring were shown to have substantial potential to affect subsequent analyses because numerous events occurred shortly after HIV diagnosis, those with and without CD4 cell counts at HIV diagnosis differed, loss to follow-up did not occur randomly and was associated with factors that were independently associated with progression. In addition to identification of all of these potential sources of bias, 91 days was determined as the most appropriate cut-off for the definition of CD4 cell counts at the time of HIV diagnosis and 365 days after a CD4 cell count was determined as the most appropriate cut-off for the definition of temporary loss to follow-up.

The analysis of late HIV diagnosis and consequent mortality, described in chapter 6, provided the highest impact public health messages from this thesis. The results demonstrated that individuals diagnosed late were approximately ten times more likely to die within a year of diagnosis than comparable individuals who were not diagnosed late. Furthermore, deaths within a year of HIV diagnosis accounted for over a quarter of all deaths among HIV-infected individuals between 2001 and 2007. I concluded that a reduction in late diagnoses could prevent substantial mortality and the literature suggests that this would be accompanied by a reduction in transmission. Factors associated with increased risk of late diagnosis and consequent mortality were identified,

indicating groups that should be offered targeted HIV testing. Due to the morbidity, mortality, financial cost and transmission potential associated with late diagnoses, guidelines in E,W&NI now recommend consideration of opportunistic screening in a variety of healthcare settings and operational research is addressing the geographical targeting of increased HIV testing in high prevalence areas. These will need to consider acceptable methods to increase demand for HIV testing among target groups as well as methods to improve the accessibility of HIV tests from a variety of health services. Late diagnosis is now monitored and reported annually by the HPA at the national and Primary Care Trust level to inform public health action. Late diagnosis is also now a priority indicator for routine monitoring at the European level by the European Centre for Disease Control.

Chapter 7 demonstrates the potential of analyses of longitudinal HIV surveillance data to describe the progression of infection among individuals to either AIDS or death as if they were being monitored as a prospective cohort. I described the absolute and proportional person years of follow-up at different levels of risk determined by CD4 cell count categories to show that the absolute time at high risk changed little over time despite marked declines in the proportion of time at high risk. The incidence of AIDS and death declined markedly during the early era of combination therapy but more slowly subsequently reflecting the immunological improvements in the population and the population effect of combination therapy. However, the incidence of AIDS and death were both continually high at low CD4 cell counts throughout the study period. AIDS within the first six months of HIV diagnosis accounted for

half of the total number of AIDS, and deaths within the first six months of HIV diagnosis accounted for one in seven of the total number of deaths. The incidence of these events during the first six months after HIV diagnosis was significantly higher than during subsequent follow-up. I believe this analysis was also the first to provide robust evidence of the effect of ethnicity on the incidence of AIDS and death after HIV diagnosis in E,W&NI. All ethnic groups had similar mortality rates but AIDS incidence was higher among black Africans, Indian/Pakistani/Bangladeshi individuals and other Asians than white individuals. This requires further investigation before the public health implications can be determined but this may depend on the improvement of surveillance systems to capture specific AIDS-defining illnesses (particularly after the time of HIV diagnosis). Heterosexual women diagnosed antenatally had lower mortality than other risk groups which may be an added benefit to the reduced transmission of infection from mother to infant and may be further strong evidence for the expansion of HIV testing among relatively low-risk populations. The disproportionately high mortality among IDUs apparent from these analyses has long been acknowledged but is not often prioritised for tertiary prevention. The routine dissemination of incidence rates in surveillance reports may highlight these outcomes and encourage public health action.

Chapters 8 and 9 focussed on utilisation of dates of starting ART that had been recently included in SOPHID. This was an essential addition to the data dictionary because it could be used, among other things, to determine whether follow-up was while the patient was on or off therapy. Chapter 8 investigated whether the date of starting ART could be reliably estimated for individuals with

only an approximate date of starting ART. CD4 cell count measurements were the only date-specific source of information available from surveillance data that were plausibly related to patients' initiation of ART. Unfortunately, none of the algorithms investigated could substantially enhance the surveillance data, either because the natural and measurement variability in CD4 cell counts was too great or because other unmeasured factors were stronger factors associated with initiation of therapy. The analysis showed that some individuals experienced a CD4 cell count below the threshold recommended by guidelines, or AIDS, before the time of startART. There may be a delay between the measurement of the CD4 cell count that stimulates the recommendation of ART and the actual date of initiation, which should be further investigated to determine any public health implications, causes and necessary remedial actions.

Chapter 9 describes the rate of increase of CD4 cell counts after starting ART, which is of interest as a predictor of morbidity and mortality among HIV-infected individuals on ART. Specific studies and cohort analyses have quantified immunological responses to ART and indicated CD4 cell counts that may be attainable for HIV-infected patients. However, these analyses of the integrated dataset show the population-level immunological response to ART in the E,W&NI and associated factors. The analysis should be repeated on more recent data with more complete information of the exact date of starting ART collected prospectively in order to clarify the findings from the first three months after starting ART. This is because of the likely bias introduced by either using the approximate date of starting ART or excluding those with only an

approximate date of starting ART. The slower rates of increase in CD4 cell counts among heterosexual men, heterosexual women (not those diagnosed antenatally) and IDU in comparison to MSM during the first three months may suggest additional adherence support needs among the former groups but requires clarification before public health interventions are initiated. Heterosexual men and IDU also had slower rates of increase in CD4 cell counts than MSM after the first three months of ART, which may again reflect adherence support needs with less likelihood of bias from approximate dates of starting ART. Data indicated that CD4 cell counts tended to continue to increase for up to eight years after starting ART although this was only statistically different for the first three years. Individuals starting ART with higher CD4 cell counts had slower rates of increase but this may have been outweighed by a reduced risk of AIDS and death. Determination of rates of increase of CD4 cell counts after starting ART in E,W&NI may provide benchmarks to help identify patients with slower immunological responses, for clinical action, or those with greater immunological responses for further public health research.

In summary, I have conducted a number of analyses that may help to inform public health responses to the HIV epidemic. These either have been, or I hope, will be integrated into national surveillance and public health processes. Lessons have been learnt for surveillance methodology with regards sharing information and ensuring that data are representative of the whole population.

10.2 Suggestions for further work

Firstly, further work should focus on issues that are likely to further validate the data to increase the reliability of results. This could include follow-up investigation of samples of i) non-linked records, ii) multiply-linked records, and iii) inconsistent records for a deeper understanding of data collection and reporting. This could also include further investigation of diagnosis before first presentation in the UK, the under-reporting of AIDS, and detection of seroconversion at HIV diagnosis. Furthermore, record linkage to the National Study of HIV in Pregnancy and Childhood dataset could assess and maximise the completeness of data on pregnancies among HIV-infected women.

A few of the analyses I would like to conduct on the data include: i) an investigation of individuals with discordant virological and immunological responses to ART including the identification of cases and controls for more in-depth study; ii) analyses of factors associated with different causes of death; iii) a repeat of some of these analyses on randomised control trial data or clinical cohort data to investigate whether similar results are apparent from datasets that do not require integration; iv) investigation with clinical case note review to assess whether multiple CD4 cell counts prior to treatment improve clinical decision-making due to more accurate determination of immunosuppression; v) survival analysis of the integrated dataset for follow-up time post diagnosis and pre-ART and separately for follow-up time post ART; vi) logistic regression of factors associated with virological response to ART.

Finally, I would like to investigate the possibility of conducting nested research studies on samples of patients drawn from the national cohort. This could utilise the national surveillance dataset as a patient-specific denominator for extrapolation of results to the national level for enhanced policy-making.

10.3 Recommendations for surveillance

The limitations of the work for this thesis indicated some recommendations for surveillance. The first of these was to improve routine data validation as records are added to the databases by record linkage. This should maximise the accuracy and internal consistency of patient records and include validation of longitudinal data for coherence, including rates of change of CD4 cell counts. Secondly, to investigate and address the fact that pregnancy status and AIDS at HIV diagnosis have been under-reported since 2000 if no clinician HIV report form is received.

Most importantly, I believe that we should enhance the monitoring of outcomes of HIV treatment and care in E,W&NI. Firstly, this would involve improvement of AIDS surveillance to capture first and subsequent AIDS diagnoses and specific AIDS-defining illnesses post HIV diagnosis. This should inform whether certain population groups are at disproportionate risk of, or certain factors are potential factors associated with progression to, specific opportunistic infections. Secondly, continue and improve the collection of treatment start and stop dates and viral loads to routinely monitor individual-level responses to treatment for research and performance management.

Finally, I suggest that the incorporation of complex indicators, such as AIDS and death incidence and the immunological response to ART, should be included in routine surveillance analyses and reports to better inform public health action.

10.4 Closing remarks

Through studying for this thesis, I developed an in-depth knowledge of HIV epidemiology, HIV datasets and mechanisms by which epidemiological data are collected and processed including a thorough understanding of the literature. I learnt more about potential sources of bias and methods to characterise bias in order to better interpret results of analyses. I used statistical techniques on longitudinal patient data that can be applied to other epidemiological data from either surveillance or research. I will be able to apply this knowledge to the development of surveillance systems and improvement of public health information to help address disease monitoring.

Appendix A: Supplementary background information

A.1 Legislation protecting the confidentiality of patient information

Voluntary and 'pseudonymised' HIV surveillance is included with other communicable disease surveillance as a part of the cover afforded to the Health Protection Agency by the Secretary of State granted under Section 251 regulations of the NHS Act 2006 (previously Section 60 of the Health and Social Care Act 2001 [in conjunction with Statutory Instrument 1438 approved in June 2002])[547]. Approvals made under these powers for the provision of data to support essential NHS services are overseen by the independent Ethics and Confidentiality Committee (previously Patient Information Advisory Group), which is a subcommittee of the National Information Governance Board (NIGB). Communicable disease surveillance, including HIV surveillance, has approval (renewed each year) under Section 251 with NIGB approval. Even with these regulations (which cover potential litigation around the common law duty of confidence) the HPA is still subject to all the requirements of the Data Protection Act.

A.2 AIDS Case Definition

Table A.1. Clinical diagnoses forming the 1993 European AIDS Case Definition[+].

AIDS indicator disease (definite or presumed)		At date of AIDS indicator disease		
		No HIV test*	HIV test positive[$]	Never tested positive and currently negative[*]
Candidiasis: trachea, bronchi, lungs	Def	✔	✔	✔ with CD4<400
Candidiasis: oesophageal	Def	✔	✔	✔ with CD4<400
	Pres		✔	
Cervical carcinoma: invasive	Def		✔ since 93	
Coccidioidomycosis: extrapulmonary	Def		✔	
Cryptococcosis: extrapulmonary	Def	✔	✔	✔ with CD4<400
Cryptosporidiosis: with diarrhoea>1 month		✔	✔	✔ with CD4<400
Cytomegalovirus (CMV) retinitis	Def	✔	✔	✔ with CD4<400
	Pres		✔	
CMV disease: not in liver spleen or nodes	Def	✔	✔	✔ with CD4<400
Encephalopathy (dementia) due to HIV	Def		✔	
Herpes simplex: ulcers for >1month or bronchitis, pneumonitis, oesophagitis	Def	✔	✔	✔ with CD4<400
Histoplasmosis: disseminated/extrapulmonary	Def		✔	
Isosporiasis: with diarrhoea >1month	Def		✔	
Kaposi's sarcoma	Def	✔ age<60	✔	✔ with CD4<400,age<60
	Pres		✔	
Lymphoma: Burkitt's or equivalent	Def		✔	
Lymphoma: immunoblastic or equivalent	Def		✔	
Lymphoma: primary in brain	Def	✔ age<60	✔	✔ with CD4<400,age<60
Mycobacterium avium: extrapulmonary	Def	✔	✔	✔ with CD4<400
	Pres		✔	
Mycobacterium tuberculosis: pulmonary	Def		✔ since 93	
	Pres		✔ *since 93*	
Mycobacterium tuberculosis: extrapulmonary	Def	✔	✔	
	Pres		✔	
Mycobacterium: other, unidentified or disseminated	Def		✔	
	Pres		✔	
Pneumocystis pneumonia	Def	✔	✔	✔
	Pres		✔	
Pneumonia, recurrent within 12 months	Def		✔ since 93	
	Pres		✔ *since 93*	
Progressive multifocal leukoencephalopathy	Def	✔	✔	✔ with CD4<400
Salmonella septicaemia, recurrent	Def		✔	
Toxoplasmosis of the brain	Def	✔	✔	✔ with CD4<400
	Pres		✔	
Wasting syndrome due to HIV	Def		✔	

[+] In those aged > 13 at date of diagnosis

* In absence of other non-HIV cause of immunodeficiency e.g. immunosuppressive therapy, other immunodepressive disease or genetic or acquired immunodeficiency syndrome atypical of HIV.

[$] For indicator disease diagnoses requiring HIV positivity for the case definition to be satisfied, e.g. Presumptive oesophageal candidiasis, the date of the HIV test should be known, and no later than the date of indicator disease diagnosis. If patient is still being treated for tuberculosis (TB) at time of HIV test, can be regarded as still having TB.

A.3 UK postcodes

UK postcodes are alphanumeric codes between five and eight characters long including a single space. This space separates the part of the postcode that identifies the postal district from the part of the postcode that identifies the group of residences (e.g. the postcode for The HPA Centre for Infections is NW9 5EQ: 'NW' is the postal area of 'North West London', 'NW9' is the postal district, 'NW9 5' is the postal sector and 'NW9 5E' is the postal sector+). A house or flat name or number is usually required to identify a specific place of residence. Postcodes are required to allocate individuals to a NHS Primary care Trust (PCT) of residence for planning, prevention and commissioning purposes. Postcodes are requested because many data reporters do not have the facilities to allocate the PCT of residence themselves. However, once the postcode information has been used to allocate a PCT of residence and used to link records from the same individual, it is encrypted along with an anonymous record number and stored separately on a secure server in a locked room. This means that it can be recalled if necessary but that the databases remain 'pseudonymised' (for example, postcodes were recently accessed so that the place of residence for every record in the database could be reassigned following changes in the structure of the NHS).

A.4 Soundex coding

The first letter of a surname is always retained and followed by three digits according to eight rules:

1. The first letter of the surname is always retained, followed by three digits selected as follows:

   A, E, I, O, U, Y, H and W are not given a code number.

B, F, P, V	Code 1
C, G, J, K, Q, S, X, Z	Code 2
D,T	Code 3
L	Code 4
M, N	Code 5
R	Code 6

2. Consonants after the initial letter are coded to the numbers above in the order in which they occur.

   e.g. Holmes H-452 Adomomi A-355

3. The code never has more than three digits, so further consonants in long names are ignored. Zeros are used for the remaining digits in short names.

   e.g. Vonderlehr V-536 Bailey B-400 Shaw S-000

4. Double consonants and adjacent consonants from the same letter group are ignored.

e.g. Ball B-400 Jackson J-250

5. A consonant immediately following a surname initial from the same letter group is ignored.

   e.g. Scanlon S-545

6. Abbreviated prefixes are coded as if they were spelt out in full.

   e.g. McIlhaney = Macilhaney M-245 St John = Saint John S-532

7. An apostrophe is ignored and the whole of double-barrelled names are coded as a single name.

   e.g. O'Neill O-540 El Eryan E-465 King-Smith K-525

8. Consonants from the same letter group separated only by W or H are only coded once.

   e.g. Booth-Davis = B-312

A.5 Algorithm for the hierarchical categorisation of exposures

1 sex between men

2 IDU

3 transfer of blood, tissue or blood products such as blood factor

- 3.1 exposure abroad
- 3.2 exposure in the UK

4 heterosexual sex

- 4.1 exposure to 'high-risk' partner(s) i.e. to partner presumed infected through sex between men, IDU, or transfer of blood, tissue or blood products
- 4.2 exposure to presumed heterosexually infected partner(s)
  - 4.2.1 exposure abroad
    - 4.2.1.1 Africa
    - 4.2.1.2 Latin America/Caribbean
    - 4.2.1.3 Asia
    - 4.2.1.4 North America
    - 4.2.1.5 Europe
    - 4.2.1.6 Australasia
    - 4.2.1.7 Country(ies) not known
  - 4.2.2 exposure in the UK
    - 4.2.2.1 to a partner infected outside Europe
    - 4.2.2.2 to a partner infected in Europe
- 4.3 partner(s) exposure category undetermined

5 occupational (healthcare), nosocomial and other blood contact

An individual who had potentially acquired infection via heterosexual sex in the UK with a partner presumed to have been heterosexually infected outside Europe would be assigned to the category numbered 4.2.2.1 in the schematic above. If that individual was thought to have acquired infection via heterosexual sex abroad with a partner presumed to have been heterosexually infected in Africa they would be assigned to the category numbered 4.2.1.1.

The hierarchy does not distinguish between risks within each category determined by factors including sexual behaviour, transmissibility of the virus, and host factors.

A.6 Development of HARS databases

At the establishment of AIDS reporting in 1982, coded data from AIDS and death reports were double entered into an in-house database and archived monthly. In 1985, with the advent of surveillance for HIV diagnoses, an EPI-INFO database was set up to collate reports of positive results from HIV tests. Data were not double entered to this database. In 1995 a unified patient-based Microsoft ACCESS ‘95 database was created in which each HIV-infected individual was recorded only once and linked to reports of HIV diagnoses and AIDS diagnoses using patient identifiers. This refined the data processing system and allowed for future volume and support of the database. In addition, individuals with an AIDS report but no HIV report could be regarded as HIV-infected because HIV was then accepted as being the cause of AIDS. Minor programming changes were made to the Microsoft ACCESS ‘95 database between 1995 and 2000 to accommodate adaptations of the reporting forms and in 2000 the database was upgraded to Microsoft ACCESS ’97. This upgrade added fields to the patient table and a form for use with the CHRs used by clinicians to report newly diagnosed HIV infections. A further upgrade from Microsoft ACCESS ’97 into Microsoft SQL Server was undertaken in January 2005 to consolidate the functionality, simplify maintenance and development, and increase the stability and responsiveness of the database.

A.7 'XLatest': the HARS archive table

Table A.2. List of fields in 'XLatest': the HARS archive table*

Field name	Data Type	Description	Field name	Data Type	Description
Patient ID	Number	Internal ID of patient in its source dataset	Viral Load	Number	Viral load (at time of clinician's report of new diagnosis)
Soundex	Text	Soundex of surname	Date VLoad	Date/Time	Date Viral Load assessed
Init	Text	Initials of patient	Has AIDS	Number	Has AIDS = 1
DOB	Date/Time	Date of Birth	Date AIDS Diag	Date/Time	Date AIDS Diagnosed
Sex	Number	Sex	Date AIDS Rep	Date/Time	Date of first AIDS Report
HIV Type	Number	HIV 1, 2, ...	Age AIDS	Number	Age at AIDS
Infection Route	Number	Risk group	SHA AIDS Rep	Text	Strategic health authority of AIDS report
Risk (Hetero)	Number	Heterosexual Risk Factor	Reg AIDS Rep	Text	New region of AIDS report
Exp	Number	Exposure	Reg AIDS Name	Text	Name of new region of AIDS report
ExpC	Number	Exposure of Contact	Was Reg AIDS Rep	Text	Pre-PCT region of AIDS report
Inf Loc	Number	Infection Location	Country AIDS	Text	Country AIDS
InfC Loc	Number	Infection Location of Contact	Clin	Number	Clinical Diagnosis
Date Last Neg	Date/Time	Date Last Negative	WHO Diagnosis 1	Number	WHO coding of Diagnosis
Date Pos	Date/Time	Date first HIV+	WHO Diagnosis 2	Number	WHO coding of Diagnosis
Ethnic Gp	Number	Ethnic Group	WHO Diagnosis 3	Number	WHO coding of Diagnosis
Date LKA	Date/Time	Date Last known alive	WHO Diagnosis 4	Number	WHO coding of Diagnosis
Source LKA	Number	Source of Last known alive	Is Dead	Number	Is Dead = 1
Resident	Number	UK Resident=1	Date Death Rep	Date/Time	Date of Death Report
Has Lab	Number	Has HIV lab report = 1	Date Death	Date/Time	Date of Death
Date Lab Rep	Date/Time	Date of first HIV (lab) Report	Age Death	Number	Age at Death
Date Lab Spec	Date/Time	Date of specimen of lab report	Pre-AIDS Death	Number	Pre-AIDS death
Age Lab	Number	Age at HIV (lab) report	Reg Death	Text	New region of death
SHA Lab Rep	Text	Strategic health authority of HIV (lab) report	Reg Death Name	Text	Name of (new) region of death
Reg Lab Rep	Text	New region of HIV (lab) report	Was Reg Death	Text	Pre-PCT region of death
Reg Lab Name	Text	Name of new region of HIV (lab) report	Status of Death	Number	Status of Death information
Was Reg Lab Rep	Text	Pre-PCT region of HIV (lab) report	Child Inf Status	Number	Infection status for a child
Country Lab	Text	Country HIV (lab) report	Year earliest Event	Date/Time	Date of first event
Has CHR	Number	Has Clinician's report = 1	Year earliest Report	Date/Time	Date of first report
Date CHR Spec	Date/Time	Date of CHR diagnosis (date of specimen on which clinician's diagnosis based)	Earliest Organisation	Text	Organisation ID of earliest report - currently only available for HARS data
Date CHR Rep	Date/Time	Date of first clinician's report	Earliest SHA	Text	First strategic health authority of report (code)
Age CHR	Number	Age at CHR	Earliest SHA Name	Text	Name of first strategic health authority of report
SHA CHR Rep	Text	Strategic health authority of clinician's report	Earliest Reg	Text	First region of report - coded to new regions
Reg CHR Rep	Text	New region of CHR report	Earliest Reg Name	Text	Name of first (new) region of report
Reg CHR Name	Text	Name of new region of CHR report	Age at earliest Event	Number	Age at earliest event
Was Reg CHR Rep	Text	Pre-PCT region of CHR report	Earliest Event Code	Number	1 - HIV lab report; 2 - CHR report; 3 - AIDS report; 4 - Death; 9 - unknown.
Country CHR	Text	Country CHR report	Latest Date	Date/Time	Date of most recent event
Cbirth	Number	Country of Birth (only from CHR)	Latest Rep	Date/Time	Date of most recent report
Year of Arrival	Number	(For patients not born in UK) year of arrival as recorded on CHR report	Latest Reg	Text	Last region of report - coded to new regions
CD4 count	Number	CD4 count (at time of clinician's report)	Latest Reg Name	Text	Name of last (new) region of report
Date CD4	Date/Time	Date of CD4 count at HIV diagnosis			

* includes data from Scotland and from the Institute of Child Health.

A.8 Methodology for deduplication of HARS records

Routine deduplication occurred when new reports are added to the database. Only records that had at least a fully matched soundex or date of birth as well as the same sex and scores above 60 were considered as possible duplicates and were assessed subjectively by users. The weighting score was loosely based upon the inverse frequency at which the data occurred in the database in December 1995 (Table A.3). Exact matches were given more weighting than fuzzy matches (those with some matching and some discordant data). Missing values had no effect on the scoring system.

Table A.3. Multiplicative scoring system for identifying possible duplicate reports

Variable	Match	Weighting	Number of different values*	Range of frequency of values*
Date of birth	Exact	25	11,212	1 – 14
	Fuzzy	2 if only the years match, 7 if 2 of day, month or year match		
Soundex	Exact	10	2,997	1 – 262
	Fuzzy	5, 7.5 or 8.3 depending on whether first, second or third digit is the first discordant one (the surname initial must agree)		
Sex		2 if both male, 4 if both female, 0.5 if discordant	2	3,231 – 22,512
First initial	Exact	2 (1st initial only)	26	1 – 1,639
Route of infection	Exact	3	8	27 – 16,696
Ethnic group	Exact	4	7	171 – 12,442

* Frequency at which variables occurred among cumulative new diagnoses in England and Wales reported to end of December 1995 as defined by year of earliest report. December 1995 was used to reflect the integration of the HIV and AIDS patient databases.

The maximum score of 6,000 could be obtained if all variables matched, whereas a score of 20 would result from a match between two records that had the same soundex, male sex and same route of infection (would therefore not be assessed).

Supporting evidence for a true match included the following: the soundex beginning with infrequent letters (I, Q, U, X, Y or Z); matching dates of death; matching GUM clinic identifiers; matching locations of diagnosis (town or city outside London, hospital or clinic within London).

Information from possible duplicate records is presented on the screen to help users decide whether they should be linked. The data shown are: local patient ID number, soundex, initials, date of birth, sex, source hospital of HIV diagnosis, source hospital of AIDS diagnosis, date of death, likelihood of match score, route of infection, country of infection, ethnicity, earliest date of diagnosis in the UK. Duplicate scores are presented in descending order with the highest score representing the most likely match. Further details on previously entered records are viewed for clarification and, if necessary, the original forms are examined for further notes. The user entering the data must decide whether to link the new report to any of the previously entered records. A different user must confirm this deduplication before it is stored as such in the database.

In addition to the routine deduplication that occurs when new reports are added to the database, quarterly deduplication is also carried out on the whole database to identify duplicate records that have not previously been detected.

Dates of birth, sex and forename initial are used to identify potential duplicate records from the same individual that have been reported with different soundex codes. The process is similar to deduplication as described above. Data presented to aid the decision are: local patient ID number, soundex, initials, date of birth, sex, health authority of HIV diagnosis, date of diagnosis, date of last negative test, route of infection, country of infection, ethnicity and any comments.

A.9 Methodology for deduplication of CD4 Surveillance records

Patient records are annually deduplicated such that all CD4 counts from duplicate records are attributed to a master record. Information such as soundex code, date of birth, sex, laboratory location, and local patient ID number are used to search for matches within the database. Information not used in the searching process such as initials, date of death and CD4 counts can be assessed to determine which of the possible matches should be linked. All punctuation is removed from patient identifiers to ensure comparability.

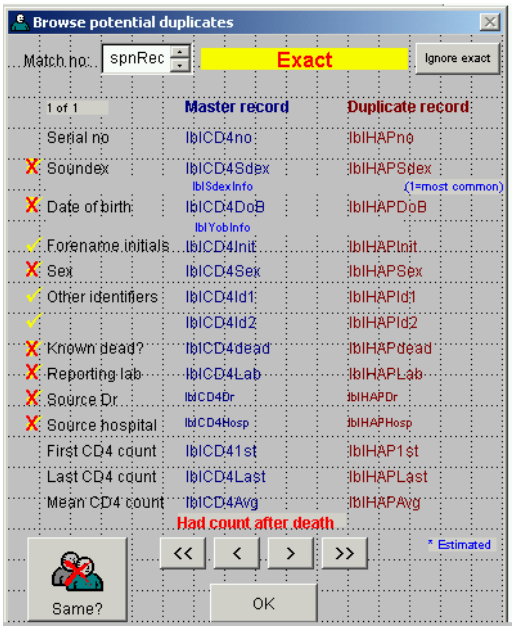
A.9.1 Deduplication process before 2005

The CD4 Surveillance application provided options of search modes (Figure A.1a and Table A.4), which when run, searched the database and presented the number of matches found, flagging those that were selected as true matches due to additional matching information. All of the potential duplicates in each search mode were stored in a temporary dataset and had to be browsed individually (Figure A.1b). The user had to decide whether to retain the decision made by the program or change that decision. Once all of the potential duplicates had been assessed the user updated the database with the decisions made. The first reported patient record was considered a ‘master record’ and the CD4 count data of its duplicates were transferred to this record. The duplicates were marked as ‘not genuine’ patients and were also marked with the record number of their master records. Patient information, such as soundex code, date of birth, sex and date of death that were missing from the master record were transferred from the duplicate record.

Figure A.1. Deduplication process a) search modes b) decision-making

a)

b)



Thousands of potential duplicate records were individually checked each year to ascertain which records were to be deduplicated and which should not. Only 3.1% (572/18,284) of the potential duplicates identified in April 2002 were considered to be true matches.

Table A.4. Results of the deduplication process, April 2002

Soundex	Date of birth	Patient ID number	Possible duplicates	Flagged as real	Deduplicated
exact	exact	ignore	658	580	294
fuzzy*	exact	ignore	4,261	286	68
exact	fuzzy*	ignore	5,831	225	21
ignore	ignore	primary = secondary	208	1	179
ignore	ignore	fuzzy primary*	-	-	-
ignore	ignore	fuzzy secondary*	7,296	0	10
Total			**18,284**	**1,092**	**572**

* fuzzy matches are those with slightly differing soundex codes or dates of birth

A.9.2 Deduplication process since 2005

An algorithm was developed to improve the efficiency of the deduplication process based on the algorithm developed for record linkage between the CD4 Surveillance and SOPHID databases (Section 3.6). This process employed a hierarchical algorithm of queries to extract potential matches according to certain criteria (Table A.5). Some of the database queries behind this process were split even further to allow more specific subsets of the matches to be linked without manual inspection. This process was a successful development of the earlier application because 80% (7,561/9,433) of the potential duplicates identified were considered to be true matches and it linked many patient records that were missing soundex codes in CD4 Surveillance.

Matches 4-15 were not repeated after 2005 because they were not greatly productive and subsequent matches were expected to pick up any valid links.

Table A.5. Results of the hierarchical deduplication process used in 2005 (records deduplicated sequentially and therefore not considered in subsequent matches)

Level	Local patient ID number	Laboratory	Date of birth	Soundex	Sex	Matches found	Deduplicated
1	✓		✓	✓	✓	27	26
2	✓		✓	✓		45	45
3	✓		✓			755	752
4	✓			✓		88	16
5	✓[1]					322	33
6	minus 1st character	✓				114	94
7	minus 1st character		✓			31	31
8	minus 1st character			✓		4	1
9	minus 1st character					819	23
10	minus first 2 characters	✓				122	46
11	minus first 2 characters		✓			8	8
12	minus first 2 characters			✓		6	4
13	minus first 3 characters	✓				73	0
14	minus first 3 characters		✓			0	0
15	minus first 3 characters			✓		0	0
16	excluding any zeros		✓			182	182
17	substring of six[2]		✓			1,831	1,686
18	substring of five[2]		✓			306	242
19		✓	✓	✓	✓	2,748	2,748
20		✓	✓	✓		118	116
21			✓	✓	✓	1,004	971
22			✓	✓		196	167
23		✓	✓		✓	18	18
24			✓	minus last character		515	314
25		✓		✓[3]	✓	7	7
26			Americanised[4]	✓		94	31
Total						**9,433**	**7,561**

[1] local patient ID number must be longer than four characters

[2] an exact match between any six/five consecutive characters of the local patient ID number

[3] and same firstname initial

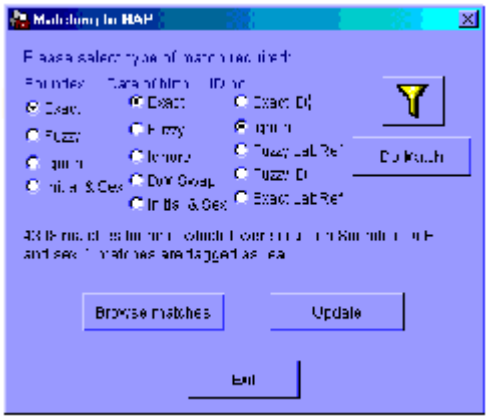
[4] switched day and month of birth

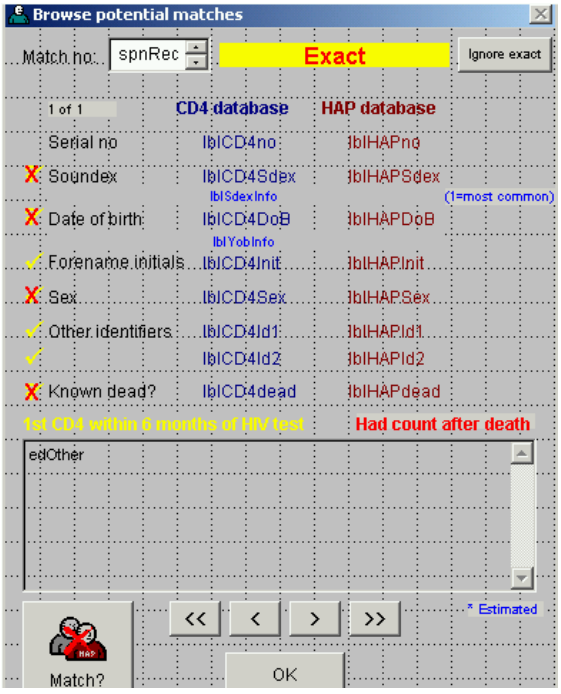
A.10 Methodology for record linkage between HARS and CD4 Surveillance

A.10.1 Record linkage to HARS before 2004

Record linkage was carried out in a similar manner to the deduplication process in CD4 Surveillance before 2005 (Appendix A.9.1). However, search modes (Figure A.2a and Table A.6), allowed the user to browse a temporary dataset of matches between the HARS and CD4 database. The user was presented with an abbreviated set of demographics (Figure A.2b) and other information from each database and had to select true matches as in the deduplication process. The user's decisions were aided by the frequency of the soundex code and date of birth from auxiliary tables. The results of the decision-making process were used to update records in CD4 Surveillance with the patient ID number from HARS.

Figure A.2. Record linkage process a) search modes b) decision-making





Of the 12,346 records that were considered as possible matches between the HARS (March 2002 archive) and CD4 databases, 2,636 (21%) were judged to be true matches. The majority of these were matched with exact soundex and date of birth or with patient ID numbers.

Table A.6. Record linkage to HARS, April 2002

Soundex	Date of birth	Patient ID number	Possible matches	Flagged as true	Linked
exact	exact	ignore	1,684	1,499	1,396
ignore	ignore	exact secondary	574	0	269
ignore	ignore	exact primary	1,248	0	844
fuzzy	exact	ignore	3,132	184	37
exact	fuzzy	ignore	4,543	173	42
ignore	ignore	fuzzy primary	75,490*	-	-
ignore	ignore	fuzzy secondary	598	0	0
exact	Date/month swap	ignore	9,394*	19	-
exact soundex + initial + sex	Date/month swap	ignore	567	19	48
Total			**12,346**	**1,875**	**2,636**

* these possible matches were not assessed as the yield was judged to be too low for the time required

A.10.2 Record linkage to HARS since 2004

As with deduplication (Appendix A.9.2), an algorithm was developed to improve the efficiency of record linkage to HARS based on the algorithm developed for record linkage between the CD4 Surveillance and SOPHID databases (Section 3.6). A key change to that process was that soundex codes and dates of birth missing from reports to CD4 Surveillance but identified from record linkage to SOPHID, were included in CD4 record linkage to HARS. Results indicated that there was a high degree of record linkage between CD4 Surveillance and

SOPHID (Chapter 3), which was expected as these reports should be received concurrently for the same individuals and with the same local patient ID numbers. It was expected that greater completion of soundex codes and dates of birth in CD4 Surveillance would improve record linkage to HARS, which was more likely to have different local patient ID numbers associated with the time of diagnosis.

The hierarchical algorithm originally considered all criteria that might result in possible matches between the CD4 Surveillance and HARS databases (Table A.7). Records linked were excluded from subsequent matches. This ensured that only the strongest link would be retained if there was more than one potential match to the same record.

The results of the original process showed that there were relatively small numbers of records that matched at the lower levels and that this did not substantially improve the record linkage. Possible matches at levels 15, 16 and 19-21 (italicised) were therefore not included in subsequent record linkage as they were judged to be insufficiently specific.

Table A.7. Results of the hierarchical record linkage process used in 2004 (records linked sequentially)

Level	Date of birth	Soundex	Sex	Local patient ID number	Exposure	Ethnicity	Location	Records linked
1	✓	✓	✓	X				21,098
2	✓	✓	✓	✓				11,060
3	✓	✓	✓	substring of five[1]				2,625
4	✓	✓		✓				41
5	✓			✓				516
6		✓		✓				353
7				✓			✓	62
8	✓			substring of five[1]				248
9		✓		substring of five[1]				140
10	✓	✓			✓	✓		72
11	✓	✓			✓			27
12	✓	✓				✓		23
13	✓	✓					✓	8
14	✓	fuzzy [2]	✓		✓	✓		111
15	✓	*fuzzy [3]*	✓		✓	✓	✓	*277*
16	✓		✓		✓	✓	✓	*26*
17	✓	fuzzy [2]	✓		✓	missing		157
18	fuzzy [4]	✓	✓		✓			77
19	*fuzzy [5]*	✓	✓					*45*
20	*fuzzy [6]*	✓	✓		✓			*1,868*
21	*fuzzy [6]*	✓	✓			✓		*141*
Total								**38,975**

[1] an exact match between any five consecutive characters of the local patient ID number
[2] the first three characters of the soundex code are the same
[3] the first character of the soundex code is the same or the numeric characters of the soundex code are the same
[4] Americanised: switched day and month of birth
[5] flipped day or month or year (e.g. 12 versus 21)
[6] same day and year but different year, or same month and year but different day or same day and month but different year

A.11 Areas of residence reported to SOPHID

The NHS is organised into regional and local organisations with responsibility for the provision of healthcare for their residents. Data were first collected for SOPHID in 1996 (for 1995) when 97 new District Health Authorities for England had just been created from the Area Health Authorities for England and Wales. These existed with minor boundary changes until 2001 when 303 PCTs were created. However, in 2005 the number of PCTs was reduced and after a period of change, 152 PCTs have remained since 1st October 2006. About 70% of these PCTs are co-terminous with local authorities, which are responsible for the provision of social care. In Wales, five Health Authorities were created in 1998 from the previous Area Health Authorities for England and Wales. These were reorganised in 2003 into 22 local health boards to be coterminus with local authorities. Northern Ireland has four Health and Social Services Boards.

In SOPHID, Health Authority of residence was collected in 1995 and 1996 but postcodes were added to the data request in 1997 to validate the allocation of Health Authorities and allocate Health Authorities where data providers could not do so themselves. To ensure data continuity, both Health Authority and PCT of residence were requested in 2001 (since the second half of 2001 in London). Once data continuity was assured, Health Authority was no longer requested in the national survey but PCT of residence continues to be collected. Internal consistency between these residence-based data are validated on receipt of the data and checked with data providers as necessary. To provide consistent trends over time by area of residence, PCTs have been retrospectively allocated, and these data are reasonably complete from 2001 onwards.

A.12 Data collected by the SOPHID surveys.

Table A.8. List of fields collected by the SOPHID surveys (data dictionary sent to data providers).

Field Name	Details	Description	Format/Coding
SDEX	Soundex code of surname	Used to link reports from the same patient (soundexing program available on request)	e.g. A123
INITIAL	Initial of first name	Used to link reports from the same patient and to identify individuals with the same soundex and date of birth (e.g. twins)	dd/mm/yyyy
DOB	Date of birth	Used to link reports from the same patient and to calculate age	dd/mm/yyyy
SEX	Sex	Male or female	Numeric code (see field codes)
CLINID	Patient's identification code within the clinic	Used to link reports from the same patient	E.g. alpha-numerical code
PCTres	PCT of residence	Primary Care Trust Code – where postcode not available (first collected in the 2001 survey)	
LA/UAres	LA/UA of residence	Local/Unitary Authority Code – where postcode not available	
HAres	*HA of residence*	*Health Authority of residence (not collected after the 2001 national survey)*	
POSTCODE	Full (unit) postcode	Used for derivation of LA, PCT and SHA of residence and to summarise spatial distribution of residents within a health district	e.g. NW9 5EQ
SITE	Site of care	Place where patient attended for treatment	e.g. St Mary's London, or codes with allocation stated e.g. SML=St Mary's London
PEXP	Infection route	How infection was probably acquired	Numeric code (see field codes)
CLIN	Clinical stage of infection	Most advanced clinical stage patient has ever reached	Numeric code (see field codes)
DATEAIDS	Date of most recent AIDS	Date of diagnosis of most recent AIDS defining illness (this should not be defined by CD4 count)	dd/mm/yyyy
ETHN	Ethnic group	Ethnic group classification (NHS classification can be accepted)	Numeric code (see field codes)
ARV	Antiretroviral therapy	Level of antiretroviral therapy prescribed by your clinic/site **when last seen**	Numeric code (see field codes)
ARVSTART	Date of start of ARV	Date this patient first ever started a course of antiretroviral therapy – may not be HAART and may not necessarily be at your clinic/site (please estimate if exact date not known)	dd/mm/yyyy
CD4	CD4 cell count	Most recent CD4 cell count **in the survey period** (per microlitre)	Number e.g. 357
VL	Viral load	Most recent viral load **in the survey period** (number of copies per millilitre)	Number e.g. 35000
VLDATE	Date of viral load	Date of most recent viral load **in the survey period**	dd/mm/yyyy
PREVCARE	Previous HIV care at another site	Did the patient ever receive HIV treatment or care elsewhere before attending at your clinic/site? Include previous care abroad. N.B. do not include those who have had a first test and then been instantly transferred to your centre.	Y/N/NK or numeric code (see field codes)
DATEPOS	Date positive on site or date of first attendance	If PREVCARE = 'N' or 1, enter date of patient's **first positive test/diagnosis** (include 'immediate' referrals such as individuals diagnosed at a GP and referred directly to you), ***or*** If PREVCARE = 'Y' or 0, enter date of patient's **first attendance** at your clinic/site If PREVCARE = 'NK' or 9, enter the earliest date you have for that patient at your clinic/site.	dd/mm/yyyy
DLSEEN	Date patient last seen at this site or date of death in the survey period	Date patient was last seen for care within the survey period OR date of death if the person is known to have died within the period	dd/mm/yyyy

A.13 Comparison of SOPHID deduplication algorithms.

Table A.9. Comparison of SOPHID total counts of adults per year according to the traditional SOPHID deduplication algorithm and the algorithm used to create the report table.

Year	SOPHID totals based on exact soundex, date of birth and sex matching	SOPHID totals based on the record linkage algorithm shown in Table 11	Difference
1995	13,499	13,408	91
1996	13,608	13,484	124
1997	14,873	14,521	352
1998	16,560	16,126	434
1999	18,659	18,146	513
2000	21,061	20,549	512
2001	24,596	23,903	693
2002	29,338	28,490	848
2003	34,263	33,392	871
2004	39,064	38,253	811
2005	44,430	43,319	1,111
2006	48,991	48,103	888
2007	53,231	52,453	778
Overall	**89,076**	**79,596**	**9,480**

A.14 Descriptive analysis of factors associated with record linkage between HARS and SOPHID

Table A.10. Proportions of records linked between HARS and SOPHID by first letter of surname, month and day of birth

First letter of surname	Proportion of linked HARS patients (total number) n (%)	Month of birth	Proportion of linked HARS patients (total number) n (%)	Day of birth	Proportion of linked HARS patients (total number) n (%)
A	1,788 (76)	1	3,286 (77)	1	1,641 (75)
B	3,169 (79)	2	2,954 (77)	2	1,306 (78)
C	2,921 (78)	3	3,436 (80)	3	1,265 (78)
D	1,869 (78)	4	3,233 (77)	4	1,205 (80)
E	670 (77)	5	3,453 (78)	5	1,289 (77)
F	1,167 (78)	6	3,433 (78)	6	1,364 (78)
G	1,751 (78)	7	3,284 (78)	7	1,287 (78)
H	2,219 (80)	8	3,290 (78)	8	1,251 (79)
I	237 (76)	9	3,191 (77)	9	1,197 (77)
J	1,062 (75)	10	3,248 (78)	10	1,422 (79)
K	1,963 (77)	11	3,160 (78)	11	1,199 (78)
L	1,604 (77)	12	3,281 (78)	12	1,390 (78)
M	5,235 (80)			13	1,190 (79)
N	1,759 (79)			14	1,271 (78)
O	1,056 (77)			15	1,477 (78)
P	1,649 (77)			16	1,223 (78)
Q	67 (78)			17	1,272 (79)
R	1,607 (76)			18	1,221 (80)
S	3,148 (77)			19	1,237 (77)
T	1,468 (79)			20	1,353 (78)
U	101 (81)			21	1,227 (77)
V	385 (74)			22	1,262 (78)
W	1,945 (80)			23	1,290 (77)
X	22 (36)			24	1,275 (77)
Y	160 (81)			25	1,422 (80)
Z	227 (69)			26	1,213 (76)
				27	1,211 (78)
				28	1,322 (78)
				29	1,158 (78)
				30	1,158 (77)
				31	651 (80)

Appendix B: HARS reporting forms

B.1 Laboratory forms

Originally, in 1984, microbiologists reported newly diagnosed HIV infections individually on the pre-existing ‘Form 30’ for the reporting of ‘Individual infections with virus, chlamydia, rickettsia or mycoplasma’ or ‘Form 30C’ that was a ‘Composite form for viruses – not hepatitis’ (Table B.1 and Form B.1a/b). Guidance for completion was provided on an accompanying sheet. The initiation of HIV serosurveillance in March 1985 led to use of ‘Form 30A’, the ‘Composite form for HTLV 3/LAV’, which allowed the reporting of up to 9 cases per sheet and was accompanied by an HIV-specific guide for completion (Form B.1c). To prevent the segregation of HIV reporting from the standard laboratory reporting of infections, HIV was reported on ‘Form 2 – Communicable Disease Reporting' from 1989 to 1993, a form that was also used to report other viral and bacterial infections (Form B.1d). In September 1993, the first dedicated HIV form with accompanying guidance sheet was introduced and this continued with minor modifications until 2000 (variations of ‘Form L’ and ‘Form D’ – Form B.1e and B.1f). There was a brief trial of a form that included an additional sheet that microbiologists could return to clinicians for further details along with the request for the confirmatory specimen. However, this was not found to be sufficiently utilised to justify the additional complication. After the introduction of clinician reporting of newly diagnosed HIV infections in January 2000 laboratories were asked to continue reporting newly diagnosed HIV infections on ‘Form L0001’ (Form B.1g), which changed only marginally to the end of 2007 (Forms B.1h and B.1i).

Table B.1. Data collected on laboratory report forms

Data field \ Period of use	Form 30 pre 1985	Form 30 C pre 1985	Form 30 A 1985 - 1989	Form 2 [1] 1989 - 1993	Form D 93 1993 - 1995	Form D 95 1995 - 2000	Form L0001 2000 - 2001	Form L3 2001 - 2004	Form L4 2004 - 2008
Source hospital or clinic					X	X	X	X	X
Reference lab.	X	X	X		X	X		[2]	[2]
Clinician/consultant who requested the test							X	X	X
Patient soundex or surname	X	A	A or C	A or C	X	X	X	X	X
Forename	X	B or C	B or C	B or C					
Initial(s)					X	X	X	X	X
Clinic/Hospital patient identifying number			C or A	C or A	X	X	X	X	X
Laboratory patient identifying number	X		C or A	C or A	X	X	X		
Age	X	D or E	X	D or E					
Date of birth				E or D	X	X	X	X	X
Sex	X	X	X	X	X	X	X	X	X
Pregnant	X				[5]	[5]			
HIV type					X	X	X	X	X
Date of specimen	X	X	X	X	X	X	X	X	X
Ethnic group					[4]	[4]	X	X	X
Has patient been tested in this lab. before?							X		
If yes, result:							X		
If yes, when							X		
Has patient been anti-HIV tested before?					X	X		X	X
If yes, result:					X	X		X	X
If yes, date of last negative test								X	X
If yes, place of last negative test								X	X
If yes, when					X	X			
If yes, where					X	X			
Has patient been HIV-diagnosed before?	X						X	X	X
If yes, when	X						X	X	X
If yes, where							X	X	X
Details of any recent seroconversion					X	X			
Facility where specimen taken					X	X	X	X	X
Clinical features at date of specimen	X	X	X	X	X	X			
Infection contributed to death?	X			X					
Died?	X		X						
Date of death					[5]	[5]			
Main reason for request	X		X		X	X			
Exposure(s) of patient			X	X	X	X	X	X	X
Probable country(ies) of infection of patient					[4]	[4]	X [3]	X [3]	X
Exposure(s) of partner					X	X	X [3]	X [3]	X
Probable country(ies) of infection of partner							X [3]	X [3]	X
Comment:					X	X	X	X	X
Reported by	X	X	X	X	X	X	X	X	X
Date report sent	X	X	X				X	X	X
Reporting laboratory	X	X	X	X	X	X	X	X	X

A = surname or soundex of surname, B = forename, C = clinic or laboratory patient ID number, D = age, E = date or year of birth

[1] Form 2 also collected the 'Organism and subtype', 'Date of onset of illness', 'Organism identified in: Specimen type', 'Organism detected by: Isolation, Direct fluorescent microscopy, ELISA, Specific immunoglobulin, Serology: X4 rise OR single titre' but these fields were not related to HIV infections.

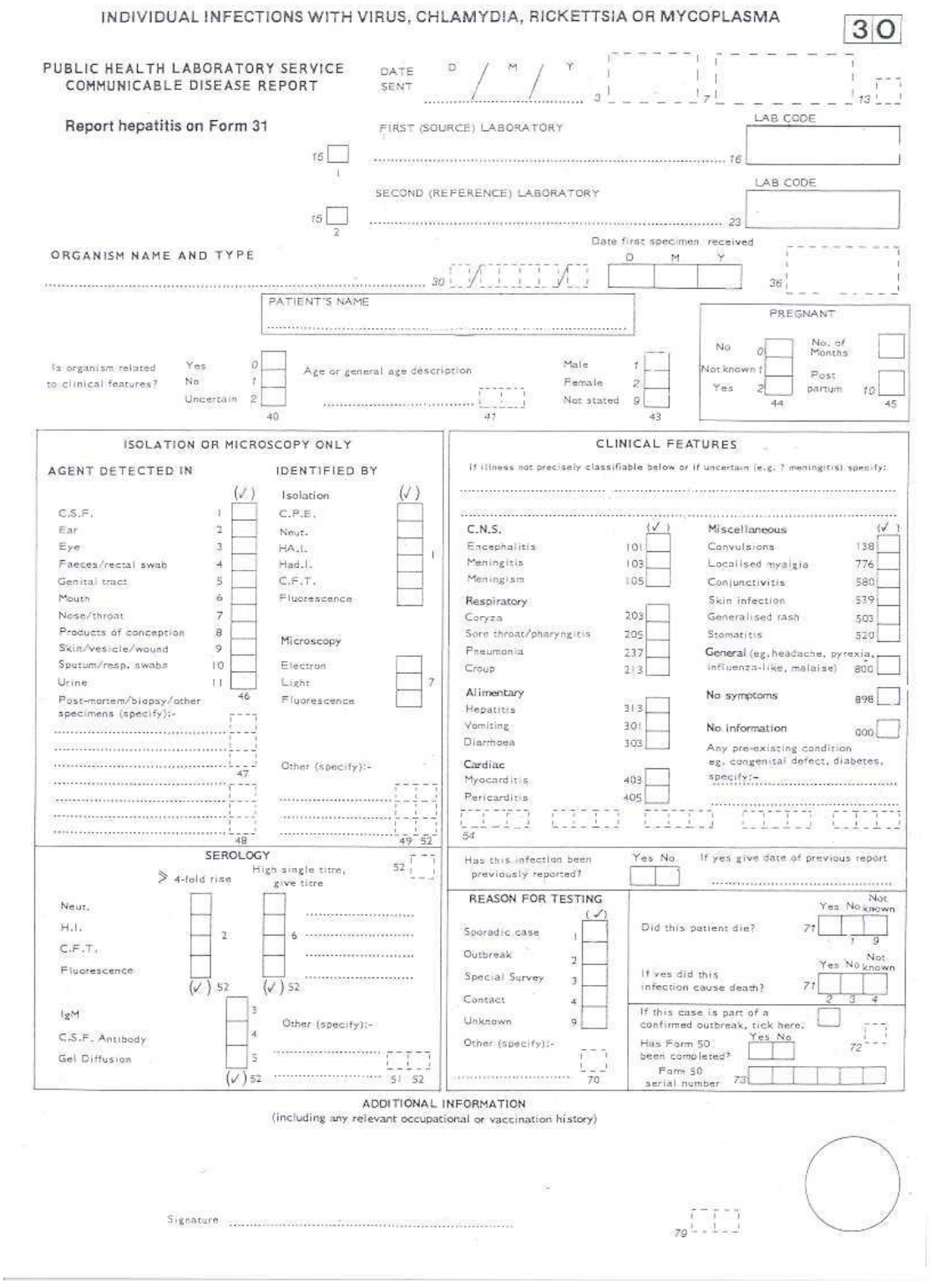
[2] only requested if HIV-2 positive

[3] only requested for individuals exposed through heterosexual intercourse

[4] only requested if infected abroad

[5] requested in comment field

Form B.1a. ‘Form 30’ for the reporting of ‘Individual infections with virus, chlamydia, rickettsia or mycoplasma’



INDIVIDUAL INFECTIONS WITH VIRUS, CHLAMYDIA, RICKETTSIA OR MYCOPLASMA

30

PUBLIC HEALTH LABORATORY SERVICE
COMMUNICABLE DISEASE REPORT

DATE SENT D / M / Y

Report hepatitis on Form 31

FIRST (SOURCE) LABORATORY — LAB CODE

SECOND (REFERENCE) LABORATORY — LAB CODE

ORGANISM NAME AND TYPE

Date first specimen received D M Y

PATIENT'S NAME

PREGNANT: No / Not known / Yes; No. of Months; Post partum

Is organism related to clinical features? Yes / No / Uncertain

Age or general age description

Male / Female / Not stated

ISOLATION OR MICROSCOPY ONLY

AGENT DETECTED IN: C.S.F., Ear, Eye, Faeces/rectal swab, Genital tract, Mouth, Nose/throat, Products of conception, Skin/vesicle/wound, Sputum/resp. swabs, Urine, Post-mortem/biopsy/other specimens (specify):-

IDENTIFIED BY: Isolation (C.P.E., Neut., HA.I., Had.I., C.F.T., Fluorescence); Microscopy (Electron, Light, Fluorescence); Other (specify):-

CLINICAL FEATURES

If illness not precisely classifiable below or if uncertain (e.g. ? meningitis) specify:

C.N.S.: Encephalitis, Meningitis, Meningism
Respiratory: Coryza, Sore throat/pharyngitis, Pneumonia, Croup
Alimentary: Hepatitis, Vomiting, Diarrhoea
Cardiac: Myocarditis, Pericarditis
Miscellaneous: Convulsions, Localised myalgia, Conjunctivitis, Skin infection, Generalised rash, Stomatitis, General (eg. headache, pyrexia, influenza-like, malaise), No symptoms, No information
Any pre-existing condition eg. congenital defect, diabetes, specify:-

SEROLOGY

≥ 4-fold rise / High single titre, give titre

Neut., H.I., C.F.T., Fluorescence
IgM, C.S.F. Antibody, Gel Diffusion
Other (specify):-

Has this infection been previously reported? Yes No — If yes give date of previous report

REASON FOR TESTING: Sporadic case, Outbreak, Special Survey, Contact, Unknown, Other (specify):-

Did this patient die? Yes No Not known

If yes did this infection cause death? Yes No Not known

If this case is part of a confirmed outbreak, tick here:

Has Form 50 been completed? Yes No

Form 50 serial number

ADDITIONAL INFORMATION
(including any relevant occupational or vaccination history)

Signature

Form B.1b. ‘Form 30C’ – ‘Composite form for viruses – not hepatitis’

COMPOSITE FORM FOR VIRUSES - NOT HEPATITIS

30C

PUBLIC HEALTH LABORATORY SERVICE
COMMUNICABLE DISEASE REPORT

DATE SENT D / M / Y

LAB CODE

Use individual form 30 for **all** single identifications, deaths, pregnant patients and interesting cases.

SOURCE LABORATORY

REFERENCE LABORATORY

ORGANISM

Use a separate form for each agent

LEAVE BLANK	NAME	AGE OR YEAR OF BIRTH	SEX M/F	ISOLATE FROM	SEROLOGY: Rising titre	SEROLOGY: IgM	SEROLOGY: Single titre	DIRECT MICRO EM IF	MAIN CLINICAL FEATURE	* OUTBREAK
	① SPEC. DATE									
	② SPEC. DATE									
	③ SPEC. DATE									
	④ SPEC. DATE									
	⑤ SPEC. DATE									
	⑥ SPEC. DATE									
	⑦ SPEC. DATE									
	⑧ SPEC. DATE									
	⑨ SPEC. DATE									

Please record additional information on reverse side of form with corresponding encircled number

*Outbreak F - family
H - hospital (name).....
I - institution (name).....
C - community

Signature.....

PUBLIC HEALTH LABORATORY SERVICE

GUIDE TO REPORTING VIRAL INFECTIONS TO CDSC

Clinical features of patients with viral infections:

all infections – please record main clinical features particularly:

conjunctivitis,
bronchiolitis, croup, pneumonia,
meningitis, encephalitis,
myocarditis, pericarditis,
rash,
and any other main clinical feature.

HIV please record on numbered line:

1. DHA of health facility from where specimen came.
2. Type of facility from where specimen came:
   GUM clinic
   outpatients
   GP
   other (specify)

3-9. Clinical features:

none,
rash,
lymphadenopathy,
candidiasis,
diarrhoea (>1 month),
weight loss (-10% baseline),
pneumonia,
Kaposi's sarcoma,
neurological disease,
fever/night sweats (>1 month),
other (please specify).

Epidemiological features of patients with viral infections:

all infections - please record if:

HIV positive,
other immunosuppression,
part of an outbreak.

also record if appropriate:

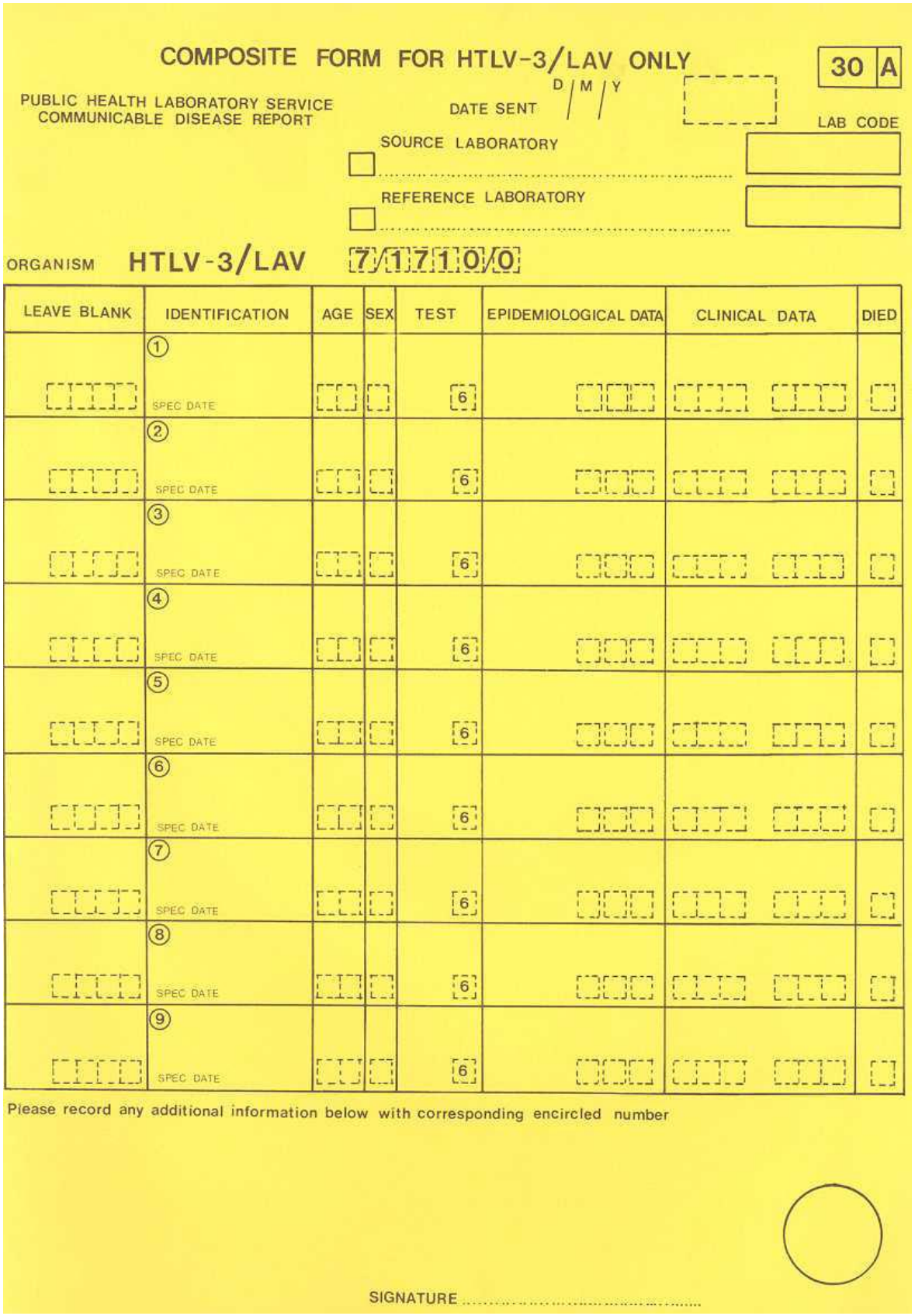
immunisation history,
recent travel abroad (state country),
congenital/infant,
pregnant,
history of contact,
occupation.

HIV please record on numbered line:

1. Risk factor(s) of patient:
   none known,
   male homosexuality or male bisexuality or homo/bisexual,
   injecting drug misuse,
   many sexual partners,
   haemophilia,
   blood/tissue recipient,
   documented HIV infection,
   lived/visited Africa,
   lived/visited central/south America,
   other (specify).
2. Risk factor(s) of contact (as in 1 above).
3. Nature of contact:
   homosexual,
   bisexual,
   intrauterine,
   other household,
   needlestick,
   blood spillage,
   bite,
   other (specify).
4. Pregnant, LMP.
5. Reason for request:
   illness,
   risk factor,
   confirmation test,
   visa screen,
   concerned patient,
   blood/tissue donor,
   insurance screen,
   other (specify).

6 – 9 Comments.

Form B.1c.‘Form 30A’, the ‘Composite form for HTLV 3/LAV’



COMPOSITE FORM FOR HTLV-3/LAV ONLY

30 A

PUBLIC HEALTH LABORATORY SERVICE
COMMUNICABLE DISEASE REPORT

DATE SENT D / M / Y

LAB CODE

SOURCE LABORATORY

REFERENCE LABORATORY

ORGANISM HTLV-3/LAV 7/1/7/1/0/0

LEAVE BLANK	IDENTIFICATION	AGE	SEX	TEST	EPIDEMIOLOGICAL DATA	CLINICAL DATA	DIED
	① SPEC DATE			6			
	② SPEC DATE			6			
	③ SPEC DATE			6			
	④ SPEC DATE			6			
	⑤ SPEC DATE			6			
	⑥ SPEC DATE			6			
	⑦ SPEC DATE			6			
	⑧ SPEC DATE			6			
	⑨ SPEC DATE			6			

Please record any additional information below with corresponding encircled number

SIGNATURE .....

Introduced Mid 85 to replace the original form 30 for individual reports of viruses, chlamydiae etc.

COMPOSITE FORM FOR HTLV 3/LAV

INSTRUCTION SHEET – PLEASE RETAIN

As far as is possible report an individual once only.

Please do not write in dotted boxes.

IDENTIFICATION: either: name
or: coded name using phonetic alpha numeric code supplied
or: laboratory/clinic number
or: leave blank

AGE: in years or give date of birth (for babies give age in months).

SEX: M, F, or NK (not known).

EPIDEMIOLOGICAL DATA: either write in information or use code number; maximum number of items is 3.

Is this patient?
1. Homosexual/bisexual.
2. Haemophiliac.
3. IV-drug user.
4. Blood transfusion recipient.
5. The spouse/baby/child/sexual or other contact (state which) of an HTLV 3 antibody positive person? If so, also state category of the contact.
6. Prostitute.
7. Central African/lived in Central Africa (state which).
8. No known risk factor identified.
9. No information.

CLINICAL DATA: either write in information or use code number; maximum number of items is 2; please avoid terms like: at risk, worried, contact, and haematological data.

1. No symptoms.
2. Lymphadenopathy/PGL.
3. Acute glandular fever.
4. AIDS/opportunist disease (infection or tumor).
5. Other (state).
6. No information.
7. Semen donor.
8. Organ donor.
9. Blood donor.

DIED: Yes, No, or NK (not known).

Please send forms to the; Director
Communicable Disease Surveillance Centre
61 Colindale Avenue
LONDON NW9 5EQ

to arrive by Monday morning post.

Form B.1d. ‘Form 2’ – ‘Communicable Disease Reporting'

PHLS COMMUNICABLE DISEASE SURVEILLANCE CENTRE — FORM 2

LABORATORY REPORTS FOR CDR

CONFIDENTIAL

Source Lab.

Reference Lab.

PATIENT IDENTIFICATION *Surname & forename (or lab. number)*	DOB *dd/mm/yy age or nk*	SEX *m/f/nk*	ORGANISM & SUBTYPE	DATE OF ONSET *dd/mm/yy*

IDENTIFIED IN:- ORGANISM DETECTED BY:-

Specimen type	Date	Isolation	Direct fluorescent microscopy	Elisa	Specific immuno-globulin	Serology *X4 rise*	Serology *single titre*
1.							
2.							
3.							

Main clinical features

1.
2.
3.
4.
5.
6.
7.
8.
9.

Infection contributed to death? Y ☐ N ☐ NK ☐

Epidemiological features

1.
2.
3.
4.
5.
6.
7.
8.
9.

Please report antimicrobial susceptibilites of bacterial isolates on reverse.

Reporting lab.

Signature

PUBLIC HEALTH LABORATORY SERVICE FORM 2

COMMUNICABLE DISEASE REPORTING TO CDSC

INSTRUCTION SHEET – PLEASE RETAIN

Forms 1 and 2 replace forms 10, 11, 20, 20M, 30, 30A, 30C, 40, 70 and fungal infections. Please continue to report hepatitis on Form 31, isolates from non-human sources on Form 41 and outbreaks on Form 50.

REPORTING INSTRUCTIONS

Report symptomatic infections only, except when others are relevant (e.g. all virus infections in pregnancy, individuals in outbreaks, typhoid, diphtheria).

Report when identification is complete. Preliminary reports of special interest or urgency may be telephoned to CDSC (01-200 6868), or reported on Form 2 marked "provisional".

Please send forms to CDSC as frequently as possible, but at least weekly.

See list for reportable organisms/syndromes.

Report the same organism, identified from one or more specimen types, on one form. Report different organisms identified from the same patient on different forms.

EXPLANATION OF HEADINGS ON FORMS

Source lab: the laboratory which received the specimen first.

Reference lab: the laboratory to which the specimen was finally referred.

Reporting lab: the laboratory completing and forwarding the form.

Patient identification: full name or Soundex code instead of name; or laboratory number of specimen.

Date of Birth (DOB): dd/mm/yy; age in d(days), w(weeks), m(months), y(years) or nk(not known).

Sex: m(male), f(female), nk(not known).

Organism: full organism name and type.

Date of onset: dd/mm/yy of onset of illness.

Specimen type(s): e.g. CSF, blood, sputum, biopsy.

Specimen date(s): dd/mm/yy specimen was collected from patient; if not known, date specimen received by first (source) laboratory.

Identification method(s): tick one or more method for each specimen type.

See the separate **'Guide to Reporting'** for

Main clinical features

Epidemiological features

Antimicrobial susceptibilities (on reverse of form): see instructions on form.

Return forms by post to:

The Director,
PHLS Communicable Disease Surveillance Centre,
61 Colindale Avenue,
London NW9 5EQ.

FORM 2 (REVERSE)

ANTIMICROBIAL SUSCEPTIBILITY OF BACTERIAL ISOLATE

If any antimicrobial susceptibilities for this organism were tested, please enter results leaving blanks as appropriate. Please record the results of β-lactamase testing when relevant, e.g. *Haemophilus influenzae, Neisseria gonorrhoeae.*

Enter results as: *S* (Susceptible)
I (Intermediate)
R (Resistant)

Amikacin	
Ampicillin/amoxycillin	
Amoxycillin/clavulanate	
Azlocillin	
Aztreonam	
Carbenicillin	
Cefotaxime	
Cefoxitin	
Ceftazidime	
Cefuroxime	
Ciprofloxacin	
Chloramphenicol	
Clindamycin	
Erythromycin	
Fusidic acid	
Gentamicin	
Imipenem	
Methicillin	
Metronidazole	
Netilmicin	
Penicillin	

Piperacillin	
Rifampicin	
Spectinomycin	
Sulphonamide	
Tetracycline	
Ticarcillin	
Tobramycin	
Trimethoprim	
Vancomycin	

OTHER ANTIMICROBIALS TESTED *Please write in*	
1.	
2.	
3.	
4.	
5.	
6.	
7.	

β-Lactamase positive	

ORGANISMS/CONDITIONS TO BE REPORTED ON FORM 2

Every organism of clinical significance isolated from **blood cultures** or metastatic infections.

Every organism causing **meningitis, encephalitis,** or **endocarditis.**

Organisms or conditions of **particular interest** (e.g. toxic shock, toxic food poisoning, osteomyelitis, septic arthritis, serious cellulitis, gangrene).

Every organism listed below:

BACTERIAL INFECTIONS

Actinomyces
Aeromonas
Anthrax
Borrelia
Brucella
Clostridium botulinum
Clostridium tetani (tetanus only)
Clostridium other (specify)
Corynebacterium diphtheria (toxigenic)
Corynebacterium other (specify)
Erysipelothrix
Gas gangrene (specify organism)
Legionella
Leptospira
Listeria
Mycobacterium
Neisseria gonorrhoeae
(extra-genital or antibiotic-resistant only)
Neisseria meningitidis
Nocardia
Ophthalmia neonatorum (by organism)
Pasteurella
Plesiomonas
Streptobacillus moniliformis
Vibrio
Yersinia

DEEP-SEATED FUNGAL INFECTIONS

Aspergillus
Candida/Torulopsis
Coccidioides
Cryptococcus
Histoplasma

HELMINTHS

Diphyllobothrium
Dracunculus
Echinococcus
Fasciola
Strongyloides
Taenia
Toxocara

PROTOZOA

Acanthamoeba
Amoebiasis (extra-intestinal or
acquired in the UK only)
Hartmanella
Leishmania
Naegleria
Pneumocystis
Toxoplasma
Trypanosoma

VIRAL AND OTHER INFECTIONS

Adenovirus (except EM)
Arbovirus
Chlamydia psittaci
Chlamydia trachomatis (non-genital only)
Cowpox
Coxiella
Coxsackie
Cytomegalovirus
EB virus (excluding uncomplicated
glandular fever)
Echovirus
Hepatitis (except A and B)
Herpes simplex (neonatal, meningitis,
encephalitis or deaths only)
HIV
HTLV
Influenza
LCM virus
Measles (SSPE, encephalitis or deaths only)
Mycoplasma
Orf/paravaccinia
Papillomavirus
Papovavirus
Parainfluenza
Parvovirus B19
Poliovirus
Polyomavirus
Rabies
Reovirus
Rickettsia
Rubella
Varicella zoster (neonatal, meningitis,
encephalitis, pneumonia, pregnant
cases or deaths only)

Form B.1e. ‘Form L 09.93’ – first dedicated laboratory HIV reporting form

PHLS COMMUNICABLE DISEASE SURVEILLANCE CENTRE
LABORATORY REPORT OF NEWLY IDENTIFIED
ANTI-HIV POSITIVE **CONFIDENTIAL**

Source Hospital | Reference Lab

Patient Soundex or Surname | Initial(s) | Clinic/Hospital No. | Date of Birth ___/___/___ | Sex | HIV ☐ 1 Comb. test ☐ 2 ☐ | Date of specimen ___/___/___ | Lab. No.

Has patient been anti-HIV tested before? ☐ Yes ☐ No ☐ Not known

If **Yes, Result:** ☐ Positive ☐ Negative When: ___/___/___ Where: .....

Please give details of any evidence of recent seroconversion .....

Facility where specimen taken

- ☐ GUM clinic
- ☐ Antenatal clinic
- ☐ Other out patient .....
- ☐ In patient
- ☐ GP
- ☐ Drug dependency unit
- ☐ Blood Transfusion Service
- ☐ Accident & Emergency
- ☐ ‘Additional’ site .....
- ☐ **Other: specify** .....
- ☐ Not known

Clinical features at date of specimen

- ☐ Not stated
- ☐ Asymptomatic
- ☐ Glandular fever like illness
- ☐ Lymphadenopathy ☐ fever (>1 mth) ☐ weight loss (>10%) ☐ diarrhoea (>1 mth)
- ☐ Pneumonia: type .....
- ☐ Candidiasis: ☐ oral ☐ oesophageal ☐ vaginal ☐ other .....
- ☐ Mycobacterial infection (inc TB): type .....
- ☐ Other opportunistic infections: specify .....
- ☐ Neurological disease: specify .....
- ☐ Kaposi's sarcoma
- ☐ **Other: specify** .....

Main reason for request

- ☐ illness
- ☐ insurance screen
- ☐ **other: specify** .....
- ☐ sexual contact of HIV +ve
- ☐ visa screen
- ☐ antenatal test
- ☐ postmortem
- ☐ blood/semen/tissue donor
- ☐ concerned patient
- ☐ not known

Exposure(s) of patient (tick all that apply and give details asked for in COMMENT)

- ☐ Homosexual intercourse
- ☐ Heterosexual intercourse:- please give details below of risk factor(s) of sexual partner(s). If exposed abroad please give details of ethnic group, possible country(ies) of infection and period(s) of exposure.
- ☐ Injecting drug use
- ☐ Child of HIV infected mother: please report **without waiting for infection to be confirmed**, and give mother's DOB and likely risk factor(s).
- ☐ Haemophilia
- ☐ Blood/tissue recipient: please state reason for treatment, country and date of treatment
- ☐ **Other:** specify .....
- ☐ Not known

COMMENT Information asked for above and/or further features of interest e.g. pregnancy (LMP), exposure through health care work or as a sex-worker, exposure abroad, bisexuality, date of death, contact known HIV infected (continue overleaf if necessary)

Patient: .....

Contact: .....

Country(ies): .....

Reported by ..... **LABORATORY** ..... FORM L 09.93

PLEASE REPORT ALL INDIVIDUALS WHEN THEY ARE FIRST IDENTIFIED AS HIV POSITIVE IN YOUR LAB, EVEN IF THERE IS A HISTORY OF THEM HAVING PREVIOUSLY BEEN FOUND POSITIVE ELSEWHERE

PLEASE REPORT EVEN IF YOU KNOW THE INDIVIDUAL HAS ALSO BEEN REPORTED TO CDSC AS HAVING AIDS.

Source Hospital: The hospital where the patient gave the specimen.

Reference Lab: Where the specimen was confirmed as anti-HIV positive.

Patient Identification: For the identification of duplicate reports of the same individual, and to link an anti-HIV positive report with a clinician's report of AIDS in the same patient, we need the surname or its "Soundex" code, and date of birth. Surnames are not stored on your computer, but patient Soundex code and initials are. Please include any other patient identification given by the clinician.

Only record **laboratory number** if this will be useful to you in identifying the patient should we contact you for further details.

HIV-1 or 2: If combined HIV 1/2 tests have been used, and the sub-type not established, please tick box for comb. test. If awaiting sub-typing result please delay reporting until information is available.

Ethnic group: This information may be available in laboratories if it is routinely collected on the hospital information system.

Evidence of seroconversion: Please record, with date of the most recent negative test if applicable, any laboratory or clinical information which indicates the period during which the patient became infected.

Clinical features at the date of specimen: This information is used to give an indication of the clinical stage at which HIV infected people are first tested. It is therefore important to distinguish those with **no symptoms** at the time of testing from those for whom symptom information is not known. We also use symptom information to indicate individuals who we would expect to have been reported by clinicians as AIDS cases.

Exposure(s) of patient: We try to get enough exposure information to allocate each individual to a category in our standard tables. If this information is incomplete or missing and you are willing for us to contact the clinician who requested the test would you please give us their name and phone number. If the information is definitely unobtainable please state why, and we will code the report appropriately. If more than one exposure applies, but you have information suggesting their relative likelihood, please give it in the space provided.

LABORATORY: Please give the name of the institution or where your laboratory is sited i.e. Borchester PHL or Royal Borsetshire - "virology" is not sufficient.

Please send the top copy of the completed HIV forms, In Confidence, to:

The Director, PHLS CDSC, 61 Colindale Avenue, London NW9 5EQ

The bottom copy is for your records.

For more forms, details of soundex coding or with any queries ring 0181-200-6868 Ext 4562 or 4563

Form B.1f. ‘Form D 03.95’ – Laboratory HIV reporting form

PHLS COMMUNICABLE DISEASE SURVEILLANCE CENTRE
LABORATORY REPORT OF NEWLY IDENTIFIED
ANTI-HIV POSITIVE CONFIDENTIAL

*Source Hospital | *Reference Lab

*Patient Soundex or Surname	Initial(s)	Clinic/Hospital No.	Date of Birth	Sex	*HIV 1 Comb. test 2	Date of specimen	Lab. No.
			___/___/___			___/___/___	

*Ethnic group (if known):

1 White | 5 Black - Caribbean | 6 Black - African | 7 Black - Other | 8 Indian/Pakistani/Bangladeshi | 4 Other/Mixed

HAS PATIENT BEEN ANTI-HIV TESTED BEFORE? Yes 2 | No 1 | Not known 1

If Yes, Result: Positive | Negative When: ___/___/___ Where: .....

*Please give details of any evidence of recent seroconversion .....

FACILITY WHERE SPECIMEN TAKEN

1 GUM clinic | 4 In patient | 7 Blood Transfusion Service | 99 Not known
2 Antenatal clinic | 5 GP | 8 Accident & Emergency
3 Other out patient..... | 6 Drug dependency unit | Other: specify .....

*CLINICAL FEATURES AT DATE OF SPECIMEN

99 Not stated | HIV related symptoms (not AIDS defining) - specify:.....
40 Asymptomatic | AIDS defining conditions - specify:.....
Other: specify.....

MAIN REASON FOR REQUEST

1 Illness/transfer of known positive | 2 Sexual contact of HIV +ve | 3 Antenatal test | 4 Blood/semen/tissue donor
5 Insurance screen | 6 Visa screen | 7 Postmortem | 8 Concerned patient
Other: specify ..... | 99 Not known

*EXPOSURE(S) OF PATIENT (TICK ALL THAT APPLY AND GIVE DETAILS ASKED FOR IN COMMENT)

- Homosexual intercourse
- Heterosexual intercourse:- please give details below of risk factor(s) of sexual partner(s). If exposed abroad please give details of ethnic group, possible country(ies) of infection and periods of exposure.
- Injecting drug use
- Child of HIV infected mother: please report without waiting for infection to be confirmed, and give mother's DOB and likely risk factor(s).
- Haemophilia
- Blood/tissue recipient: please state reason for treatment, country and date of treatment
- Other: specify .....
- Not known

COMMENT Information asked for above and/or further features of interest e.g. pregnancy (LMP), exposure through health care work or as a sex-worker, exposure abroad, bisexuality, date of death, contact known HIV infected (continue overleaf if necessary)

Patient: .....

Patient's Contact: .....

Probable Country(ies) of infection: .....

FORM D 03.95

Reported by _____ *LABORATORY _____

*SEE NOTE ON REVERSE

Form B.1g. 'Form L0001' – Laboratory HIV reporting form

Laboratory report of newly identified anti HIV positive specimen CONFIDENTIAL

CDSC use only

			week no	Serial No

Consultant/clinician who requested the test	Clinic/ward/hospital/practice result returned to

Patient Soundex or Surname	Initial(s)	Clinic/Hospital No.	Date of Birth	Sex	Date of specimen	Lab. No.
			___/___/___		___/___/___	

PLEASE CHECK BOX IF YOU DO NOT WISH US TO CONTACT THE REQUESTING CLINICIAN FOR FURTHER INFORMATION ☐

Results: HIV 1/2 specific testing not done ☐ HIV-1 Pos ☐ Neg ☐ HIV-2 Pos ☐ Neg ☐

Has the patient been tested in this laboratory before? YES ☐ NO ☐

If YES Date of most recent previous diagnosis ___/___/___ Result POS ☐ NEG ☐ INDETERMINATE ☐

Has the patient had HIV diagnosed elsewhere? YES ☐ NO ☐ NK ☐

If YES: when? (mm/yy) ___/___ where? .....

ETHNIC GROUP (if known)

☐1White ☐5Black-Caribbean ☐6Black-African ☐7Black-Other ☐8Indian/Pakistani/Bangladeshi ☐4Other/Mixed

FACILITY WHERE SPECIMEN COLLECTED

- ☐1 GUM Clinic
- ☐2 Antenatal Clinic
- ☐3 Other out Patient (specify) .....
- ☐4 In Patient
- ☐5 GP
- ☐6 Drug dependency unit
- ☐7 Blood Transfusion Service
- ☐8 Accident and Emergency
- ☐ Post mortem Specimen
- ☐ Other: specify .....

EXPOSURE(S) OF PATIENT, IF KNOWN FROM REQUEST FORM (please tick all that apply)

- ☐ Homosexual intercourse
- ☐ Heterosexual intercourse: please give risk factor(s) of partner(s) and possible countries of infection in COMMENT
- ☐ Injecting Drug Use
- ☐ Child of HIV infected mother
- ☐ Haemophilia
- ☐ Blood / tissue recipient
- ☐ Other: specify .....
- ☐ Not Known

COMMENT

Reported by:	Date	Reporting Laboratory

PHLS Communicable Disease Surveillance Centre Form L0001

PLEASE REPORT ALL INDIVIDUALS WHEN THEY ARE FIRST IDENTIFIED AS HIV POSITIVE IN YOUR LAB, EVEN IF THERE IS A HISTORY OF THEM HAVING BEEN PREVIOUSLY FOUND POSITIVE ELSEWHERE

In January 2000 there was a change in the surveillance of new diagnoses of HIV infection in England, Wales and Northern Ireland. Previously surveillance of HIV diagnoses relied solely on microbiologists reporting the HIV positives identified in their laboratory, while the clinicians treating people with HIV infection were only asked to report those who developed AIDS. The introduction of more effective antiretroviral therapy from 1995 onward led to a rapid decline in the number of AIDS cases in those treated. As a result AIDS was no longer a largely unbiased marker of disease progression and clinicians were reporting a decreasing proportion of the patients in their care. In response to this change clinicians have been asked to report all HIV infected patients at their first UK diagnosis. Because of their direct contact with patients, clinicians can provide additional information to that available in the laboratory.

Despite the change outlined above, laboratory reporting remains vital to the surveillance of HIV infection for three reasons:

- To provide continuity with the previous 15 years of reporting,
- To indicate the diagnoses for which clinicians' reports would be expected,
- To provide basic information when clinician reporting is delayed or inappropriate.

LABORATORY REPORTING OF NEW DIAGNOSES OF HIV INFECTION REMAINS ESSENTIAL TO NATIONAL SURVEILLANCE.

Because of the changed situation the laboratory report form was simplified. Reporters are asked to complete only as much of the form as they can from information available in the laboratory at the time of making the diagnosis. We do, however, ask for information which will allow us to contact the doctor who requested the test unless the laboratory reporter asks us not to do so.

Circumstances in which the reporter will be routinely contacted for further information:

- If the fields identifying the requesting doctor and the "clinic/ward/ hospital/practice result returned to" have not been completed
- If the report does not give enough information (soundex code and date of birth **or** clinic or hospital number) to identify the patient at the place where they are receiving care
- If the date of the specimen is not given
- If the route of infection is not recorded and the form asks us not to contact the clinician providing care

PLEASE SEND COMPLETED FORMS "IN STRICT MEDICAL CONFIDENCE" TO:
The Director PHLS CDSC, 61 Colindale Avenue, London NW9 5EQ
(PHLs use DX6530002 Colindale NW)
For more forms, details of soundex coding or help with any queries ring:
020-8200-6868, ext 4406, or 4455

Form B.1h. 'Form L3' – Laboratory HIV reporting form

CONFIDENTIAL
Laboratory report of new HIV diagnosis

CDSC use only

			week no	Serial No

Consultant/clinician who requested the testing	Clinic or ward and hospital/practice result returned to

Patient Soundex or Surname	Initial(s)	Clinic/Hospital No.	Date of Birth	Sex	Date of specimen	Lab. No.
			___/___/___		___/___/___	

PLEASE CHECK BOX IF YOU DO NOT WISH US TO CONTACT THE REQUESTING CLINICIAN FOR FURTHER INFORMATION ☐

esults: HIV positive – type not known ☐ HIV-1 Pos ☐ Neg ☐ HIV-2 Pos ☐ Neg ☐

If HIV-2 Positive please state confirmatory lab, or test used .....

Had patient ever had an HIV test prior to this diagnosis? YES ☐ NO ☐ NK ☐ If YES:-

Date and place of **last negative** HIV test (if any) ___/___/___ .....

Date and place of **earliest** HIV diagnosis (if any previous to this one) ___/___/___ .....

THNIC GROUP (if known)

☐1White ☐5Black-Caribbean ☐6Black-African ☐7Black-Other ☐8Indian/Pakistani/Bangladeshi

]4Other/Mixed

ACILITY WHERE SPECIMEN COLLECTED

- ☐1 GUM Clinic
- ☐2 Antenatal Clinic
- ☐3 Other out Patient (specify) .....
- ☐4 In Patient
- ☐5 GP
- ☐6 Drug dependency unit
- ☐7 Blood Transfusion Service
- ☐8 Accident and Emergency
- ☐ Post mortem Specimen
- ☐ Other: specify .....

XPOSURE(S) OF PATIENT, IF KNOWN FROM REQUEST FORM (please tick all that apply)

- ☐1 Homosexual intercourse
- ☐6 Heterosexual intercourse:please give risk factor(s) of partner(s) and possible countries of infection in COMMENT
- ☐2 Injecting Drug Use
- ☐8 Child of HIV infected mother
- ☐4 Haemophilia
- ☐5 Blood / tissue recipient
- ☐ Other: specify .....
- ☐ Not Known

CDSC use only

COMMENT		

Reported by:	Date	Reporting Laboratory

HEALTH PROTECTION AGENCY
Communicable Disease Surveillance Centre Form L3

PLEASE REPORT ALL INDIVIDUALS TO THE HIV/AIDS REPORT SECTION *(HARS)* AT CDSC WHEN THEY ARE FIRST IDENTIFIED AS HIV POSITIVE IN YOUR LAB, EVEN IF THERE IS A HISTORY OF THEM HAVING BEEN PREVIOUSLY FOUND POSITIVE ELSEWHERE

In January 2000 there was a change in the surveillance of new diagnoses of HIV infection in England, Wales and Northern Ireland. Previously surveillance of HIV diagnoses relied solely on microbiologists reporting the HIV positives identified in their laboratory, while the clinicians treating people with HIV infection were only asked to report those who developed AIDS. The introduction of more effective antiretroviral therapy from 1995 onward led to a rapid decline in the number of AIDS cases in those treated. As a result AIDS was no longer a largely unbiased marker of disease progression and clinicians were reporting a decreasing proportion of the patients in their care. In response to this change clinicians have been asked to report all HIV infected patients at their first UK diagnosis. Because of their direct contact with patients, clinicians can provide additional information to that available in the laboratory.

Despite the change outlined above, laboratory reporting remains vital to the surveillance of HIV infection for three reasons:

- To provide continuity with the previous 15 years of reporting,
- To indicate the diagnoses for which clinicians' reports would be expected,
- To provide basic information when clinician reporting is delayed or inappropriate.

LABORATORY REPORTING OF NEW DIAGNOSES OF HIV INFECTION REMAINS ESSENTIAL TO NATIONAL SURVEILLANCE.

Because of the changed situation the laboratory report form was simplified. Reporters are asked to complete only as much of the form as they can from information available in the laboratory at the time of making the diagnosis. We do, however, ask for information which will allow us to contact the doctor who requested the test unless the laboratory reporter asks us not to do so.

Circumstances in which the reporter may be contacted for further information:

- If the fields identifying the requesting doctor and the "Clinic or ward and hospital/practice result returned to" have not been completed
- If the report does not give enough information (soundex code and date of birth **or** clinic or hospital number) to identify the patient at the place where they are receiving care
- If the date of the specimen is not given
- If the route of infection is not recorded and the form asks us not to contact the clinician providing care

PLEASE SEND COMPLETED FORMS "IN STRICT MEDICAL CONFIDENCE" TO:

HEALTH PROTECTION AGENCY,
Communicable Disease Surveillance Centre,
HARS Section, 61 Colindale Avenue, London NW9 5EQ

For more forms, pre-paid labels, details of soundex coding or help with any queries ring:
020-8200-6868, ext 4406, or 4455

Form B.1i. 'Form L4' – Laboratory HIV reporting form

Communicable Disease Surveillance Centre L4

CDSC use only

			week no	Serial No

CONFIDENTIAL
Laboratory report of new HIV diagnosis

Consultant/clinician who requested the testing	Clinic or ward and hospital/practice result returned to

Patient Soundex or Surname	Initial(s)	Clinic/Hospital No.	**Date of Birth**	**Sex**	**Date of specimen**	Lab. No.
			___/___/___		___/___/___	

Ethnic group

☐1 White ☐5 Black-Caribbean ☐6 Black-African ☐7 Black-Other ☐8 Indian/Pakistani/Bangladeshi ☐4 Other/Mixed

PLEASE TICK BOX IF YOU DO NOT WISH US TO CONTACT THE REQUESTING CLINICIAN FOR FURTHER INFORMATION ☐

RESULTS

HIV positive – type not known ☐ HIV-1 Pos ☐ Neg ☐ HIV-2 Pos ☐ Neg ☐

If HIV-2 positive please state confirmatory lab, or test used .....

Had patient ever had an HIV test prior to this diagnosis? YES ☐ NO ☐ NK ☐ If YES:-

Date and place of **last negative** HIV test (if any) ___/___/___ .....

Date and place of **earliest positive** HIV diagnosis ___/___/___ .....

FACILITY WHERE SPECIMEN COLLECTED

☐1 GUM Clinic
☐2 Antenatal Clinic
☐8 Accident and Emergency
☐3 Other out Patient: *specify* .....
☐4 In Patient
☐5 GP
☐11 Prison
☐6 Drug dependency unit
☐7 Blood Transfusion Service
☐ Post mortem specimen
☐ Other: *specify* .....

EXPOSURE(S) OF PATIENT

☐6 Heterosexual intercourse (*please give risk factor(s) for partner(s) in COMMENT)*
☐1 Homosexual intercourse
☐2 Injecting Drug Use
☐8 Child of HIV infected mother
☐4 Haemophilia
☐5 Blood / tissue recipient
☐9 Not known
☐ Other: *specify* .....

Patient's probable country of infection .....

Partner's probable country of infection *(if known)* .....

COMMENT CC		

Reported by:	Date	Reporting Laboratory

PLEASE REPORT HIV POSITIVE INDIVIDUALS WHEN THEY ARE FIRST IDENTIFIED IN YOUR LAB, EVEN IF PREVIOUSLY DIAGNOSED ELSEWHERE

Laboratory reporting of new diagnoses of HIV infection is essential to national surveillance of HIV infection.

Since the beginning of the UK's HIV epidemic laboratories have been reporting diagnoses of HIV infection to CDSC. Clinician HIV reporting was introduced in 2000 to supplement laboratory reporting of HIV by collecting additional information, such as country of birth, behavioral data and treatment information which is not readily available in laboratories. Despite the introduction of clinician HIV reporting, laboratory reporting remains vital to HIV surveillance because:

- in many cases it provides the only record of a new diagnosis, particularly for diagnoses made in non GU settings such as GPs, non GU departments, prisons and at post mortem
- it provides continuity with reporting before 2000, when laboratory reporting was the only source of information on new HIV diagnoses

Reporters are asked to complete as much of the form as possible from information available in the laboratory at the time of diagnosis.

However, there are some fields that must be completed so we can determine whether we already know about a previous diagnosis in an HIV-infected individual:

- soundex code* of surname
- date of birth
- sex
- date of diagnosis (within the UK)

If any one of these is missing from a report, we cannot immediately enter the form onto the surveillance database and will have to therefore contact the clinician for the information.

Electronic reporting of new HIV diagnoses may be possible if your laboratory records are computerised.

Please contact us to discuss the possibility of electronic reporting.

PLEASE SEND COMPLETED FORMS "IN STRICT MEDICAL CONFIDENCE" TO:

HARS, CDSC, Health Protection Agency, 61 Colindale Avenue, London NW9 5EQ

***For more forms, pre-paid postage labels, details of soundex coding or help with any queries ring: 020-8200-6868 ext 4406 or 4455**

Thank you for your continuing support

B.2 AIDS forms

Table B.2. Data collected on AIDS forms

	Period of use							
Data field	**pre 1985**	**1985 - 1987**	**1987 - 1989**	**1989 - 1993**	**1993 - 1996**	**1996 - 1997**	**1997 - 1998**	**1998 - 1999**
Newly diagnosed AIDS or death without AIDS						X	X	X
Patient soundex or surname	X	X	X	X	X	X	X	X
Forename(s)	X	X	X					
Initial(s)					X	X	X	X
Date of birth	X	X	X	X	X	X	X	X
Clinic/Hospital patient identifying number						X	X	X
Name of hospital if patient hospitalised	X	X						
Location of hospital if patient hospitalised	X	X						
Sex	X	X	X	X	X	X	X	X
Age	X	X						
Occupation of patient	X		X	X	X	X	X	X
Patient's usual address		X	X	X	X	X	X	X
Town or postal district	X	X	X	X	X	X	X	X
Full postcode			X	X	X	X	X	X
Health District/Authority (if not town/postal district)			X	X	X	X	X	X
Local Authority (if not town/postal district)				X	X	X	X	X
Is patient overseas visitor intending to return home			X	X	X	X	X	X
If yes, country of residence			X	X	X	X	X	X
If immigrant, from what country		X						
Country(s) (if patient has lived abroad)			X	X	X	X	X	X
Date from (if patient has lived abroad)			X	X	X	X	X	X
Date to (if patient has lived abroad)			X	X	X	X	X	X
Country of birth (if patient has lived abroad)					X	X	X	X
Ethnic group	X	X	X	X	X	X	X	X
Country of origin (if African, Caribbean or other)	X	X						
Marital status	X	X	X	X	X	X	X	X
Sexuality	X	X	X	X	X	X	X	X
Exposure(s) of patient	X[1]	X[1]	X[2]	X[3]	X[4]	X[4]	X[4]	X[4]
Non-injecting drug use		X						
Any further details about likely exposure(s)			X	X	X	X	X	X
Sexual contacts with Americans/Caribbeans	X	X						
Date of last negative HIV test			X	X	X	X	X	X
HIV antibody positive			X	X	X	X	X	X
HIV antigen/culture positive			X	X				
Date of first positive test			X	X	X	X	X	X
Where? (first positive test)						X	X	X
HIV-2 positive					X	X	X	X
Cause of immunodeficiency if patient not HIV+			X	X	X	X	X	X
Date of AIDS diagnosis (mm/yy)			X	X				
AIDS indicator disease (table xx.xxx)	X	X	X	X	X	X	X	X
Date of onset of AIDS indicator disease	X	X	X	X	X	X	X	X
Definitive or presumptive diagnosis				X	X	X	X	X
Site of illness	X	X	X					
Method of diagnosis	X	X	X	X				
Signs and symptoms of illness	X	X	X		X			
Unusual or interesting clinical features				X				
Date of onset of signs and symptoms	X	X	X		X			
Date of first presentation for medical advice	X	X						
Presence of serious underlying condition	X	X						

Did patient donate blood in 5 years prior to illness	X	X						
Is this case related to any other known cases	X	X						
Previously, did patient have infection/vaccination	X[5]	X[5]		X[6]				
Has patient had surgery since Dx or in last 5 years	X							
Details of HTLV III serology	X							
Lowest total white blood cell count		X						
% of lymphocytes corresponding to the above		X						
% of T cells corresponding to the above		X						
CD4 count at time of AIDS diagnosis				X	X			
Date of CD4 count					X			
Other immunological test results		X	X					
Was patient on ARV before AIDS diagnosis					X	X	X	X
If yes, what was the most intensive level of ARV							X	X
What drugs were prescribed at that time								X
Was the patient still on ARV at AIDS diagnosis							X	X
Date of first starting ARV treatment (mm/yy)						X	X	X
Did patient get Pneumocystis pneumonia (PCP) prophylaxis before AIDS Dx					X	X	X	X
Date of first starting PCP prophylaxis (mm/yy)						X		X
Did patient get other prophylaxis before AIDS Dx					X	X	X	X
Has the patient died	X	X	X	X	X	X	X	X
If yes, date of death	X	X	X	X	X	X	X	X
Cause of death			X		X	X	X	X
Name and address of notifying doctor	X	X						
Name and address of notifying consultant	X	X						
Name of reporting consultant/GP			X	X	X	X	X	X
Address			X	X	X	X	X	X
Telephone number	X	X	X	X	X	X	X	X
Completed by			X	X	X	X	X	X
Position				X	X	X	X	X
Date of report		X	X	X	X	X	X	X

[1] IV drug use and which drugs; did the patient have a blood transfusion; did the patient receive blood products

[2] misused drugs by injection, had any injection abroad, had blood transfusion, had blood products, had surgical procedures, been sexually active abroad (homo/hetero), had other known exposure to a HIV+ve person, other possible exposure. For all, when and whether in the UK or abroad

[3] injected drugs for recreational use, had a blood transfusion, had blood products for a coagulation disorder, had sexual relations abroad (homo/hetero), had sexual relations with any of the following: IDU, bisexual man, person with haemophilia/coagulation disorder, person with AIDS or known to be HIV-infected. For all, when and whether in the UK or abroad

[4] sexual intercourse between men, sexual intercourse between men and women – including possible country of infection, IDU, child of HIV-infected mother, haemophilia/coagulation disorder, blood/tissue recipient – including country and reason. For all, approximate dates

[5] YES/NO/UNKNOWN. How many times? Year of last infection. For gonorrhea, syphilis, non-gonococcal urethritis, genital herpes simplex, amoebiasis, giardiasis, hepatitis A (infectious), hepatitis B (serum), vaccination against hepatitis B

[6] YES/NO. Month and year of first diagnosis. For herpes zoster, pulmonary tuberculosis, acute hepatitis B, syphilis, gonorrhoea, non-specific gonococcal infection

Form B.2a. ‘AIDS form 83-85’ – ‘Questionnaire for surveillance of A.I.D.S.

1983-85

IN CONFIDENCE VF46/20

QUESTIONNAIRE FOR SURVEILLANCE OF A.I.D.S.

COMMUNICABLE DISEASE SURVEILLANCE CENTRE
61 COLINDALE AVENUE
LONDON NW9 5EQ

1. Name of patient: .....
(or code)
(using alpha-numeric code, provided on request, to preserve confidentiality)

2. Patient's occupation: .....

3. Age: ..... Status Alive ☐ Dead ☐

4. Date of birth: ..... Date of death: .....

5. Sex: ..... Sex preference: Male ☐ Female ☐
Both ☐ Unknown ☐

6. Marital Status: Married ☐ Never married ☐ Divorced ☐
Widowed ☐ Separated ☐ Unknown ☐

7. Patient's Ethnic Origin
(Please tick (√) as many of the boxes as are applicable:-)

Caucasian ☐ African ☐ .....
(please specify country if possible)

British ☐ Caribbean ☐ .....
(please specify country if possible)

American ☐ Other ☐ .....
(please specify country if possible)

Jewish ☐

Italian ☐

8. Patient's address: (by postal district, or town) .....

9. If hospitalised, name and location of hospital: .....
.....
.....

10. TYPE OF ILLNESS (please tick √)	SITE OF ILLNESS	DATE OF DIAGNOSIS (Month, Year)	METHOD OF DIAGNOSIS Biopsy	Culture	Other (specify)
Kaposi's sarcoma	.....	.....	☐	☐	.....
Pneumocystis	.....	.....	☐	☐	.....
CMV infection	.....	.....	☐	☐	.....
Progressive herpes simplex	.....	.....	☐	☐	.....
(please specify duration and extent):					
Toxoplasmosis (CNS)	.....	.....	☐	☐	.....
Cryptococcosis "	.....	.....	☐	☐	.....
GI candidiasis (oesophageal, etc not just thrush)	.....	.....	☐	☐	.....
Other opportunistic infection	.....	.....	☐	☐	.....
(please specify)					
Other medical condition	.....	.....	☐	☐	.....
(please specify)					

11. What were the signs and symptoms of illness?
(only those attributed to illnesses listed in question 10)

		Date of onset (Month, Year)
☐	Skin lesions (specify location)	.....
☐	Lymphadenopathy (please specify location)	.....
☐	Fever	.....
☐	Dyspnoea	.....

11. Continued

Date of onset (Month, Year)

☐ Weight Loss .....

☐ Diarrhoea .....

☐ Oral thrush .....

☐ Other (please specify) .....

.....

ON WHAT DATE DID PATIENT FIRST PRESENT FOR MEDICAL ADVICE? .....

12. Does (or did) patient have any serious underlying condition (e.g. malignancy, organ transplant or immunosuppressive therapy)?

Yes ☐ No ☐ Unknown ☐

If Yes, please specify .....

.....

13. Name and address of notifying doctor: .....

.....

..... Telephone No. .....

14. Name and address of consultant: .....

.....

..... Telephone No. .....

15. Details of HTLV III serology if done .....

16. Has the patient used IV drug in the past 10 years

Yes ☐ No ☐ Unknown ☐

If Yes, please specify (e.g. heroin) .....

17. a) Did patient have a blood transfusion before this illness?

Yes ☐ No ☐

b) Did patient receive blood products before this illness?

Yes ☐ No ☐

c) DID PATIENT DONATE BLOOD IN 5 YEARS PRIOR TO ONSET OF THIS ILLNESS?

Yes ☐ No ☐

18. Is this case related to any other known cases?

Yes ☐ No ☐ Unknown ☐

If Yes, how? .....

19. Before the onset of illness, did patient ever have:-

		Yes	No	Unknown	How many times	Year of last infection
a.	Gonorrhoea	☐	☐	☐	☐	☐
b.	Syphilis	☐	☐	☐	☐	☐
c.	Non-gonococcal urethritis	☐	☐	☐	☐	☐
d.	Genital herpes simplex	☐	☐	☐	☐	☐
e.	Amoebiasis	☐	☐	☐	☐	☐
f.	Giardiasis	☐	☐	☐	☐	☐
g.	Hepatitis A (infectious)	☐	☐	☐	☐	☐
h.	Hepatitis B (serum)	☐	☐	☐	☐	☐
i.	Vaccination against hepatitis B	☐	☐	☐	☐	☐

20. Has the patient undergone any

a). Surgical procedure since diagnosis, or in the past 5 years

Yes ☐ No ☐

b) If yes, what was it's nature and where was it performed?

.....

.....

21. Sexual contact with Americans

a) Did the patient travel to the USA or Caribbean countries prior to illness?

Yes ☐ No ☐ Unknown ☐

If Yes, to what cities and when?

	Month/Year
.....	.....
.....	.....
.....	.....
.....	.....

b) Did the patient have any sexual contacts while travelling?

Yes ☐ No ☐ Unknown ☐

If Yes, please specify:-

City	Month/Year	No. of Contacts
.....	.....	.....
.....	.....	.....
.....	.....	.....
.....	.....	.....

c. Did the patient have any sexual contacts with Americans or Caribbean nationals?

Yes ☐ No ☐ Unknown ☐

City	Month/Year	No. of Contacts
.....	.....	.....
.....	.....	.....
.....	.....	.....

Thank you for your help. Please return completed questionnaires to:

The Director
Communicable Disease Surveillance Centre
61 Colindale Avenue
LONDON
NW9 5EQ

Tel: 01-200 6868

Form B.2b. ‘AIDS form 85-87’ – ‘Questionnaire for surveillance of A.I.D.S. and Kaposi’s Sarcoma

1985-87

IN CONFIDENCE

QUESTIONNAIRE FOR SURVEILLANCE OF A.I.D.S. AND KAPOSI'S SARCOMA

Communicable Disease Surveillance Centre
61 Colindale Avenue
LONDON NW9 5EQ

Identification Number ☐☐☐☐ (to be assigned)

1. Name of patient .....

2. Age ..... Status: Alive ☐ Dead ☐

3. Date of birth ..... Date of death .....

4. Sex ..... Sexual preference: Male ☐ Female ☐
Both ☐ Unknown ☐

5. Marital status: Married ☐ Never married ☐ Divorced ☐
Widowed ☐ Separated ☐ Unknown ☐

6. Patient's Ethnic Origin

Please tick (✓) as many of the boxes as are applicable:-

Caucasian ☐
British ☐
American ☐
Jewish ☐
Italian ☐

African (please specify country if possible) ☐ .....
Caribbean (please specify country if possible) ☐ .....
Other (please specify) ☐ .....

If immigrant, from what country? .....

7. Patient's address .....
.....

8. If hospitalised, name and location of hospital:-
.....
.....
.....
.....

9.

Type of Illness (please tick √)	Site of Illness	Date of Diagnosis (Month, Year)	Method of Diagnosis: Biopsy	Culture	Other (please specify)
Kaposi's sarcoma	.....	.....	☐	☐	.....
Pneumocystis	.....	.....	☐	☐	.....
CMV infection	.....	.....	☐	☐	.....
Progressive herpes simplex (please specify duration & extent)	.....	.....	☐	☐	.....
Toxoplasmosis (CNS)	.....	.....	☐	☐	.....
Cryptococcosis (CNS)	.....	.....	☐	☐	.....
GI candidiasis (oesophageal, etc., not just thrust)	.....	.....	☐	☐	.....
Other opportunistic infection (please specify)	.....	.....	☐	☐	.....
Other medical condition (please specify)	.....	.....	☐	☐	.....

10. What were the signs and symptoms of illness?
(only those attributable to illnesses listed in question 9)

		Date of onset (Month, Year)
☐	Skin lesions (please specify location)	.....
☐	Lymphadenopathy (please specify location)	.....
☐	Fever	.....
☐	Dyspnoea	.....
☐	Weight loss	.....
☐	Diarrhoea	.....

Date of onset (Month, Year)

☐ Oral thrush .....

☐ Other (please specify) .....

.....

ON WHAT DATE DID PATIENT FIRST PRESENT FOR MEDICAL ADVICE? .....

11. Does (or did) patient have any serious underlying condition (e.g. malignancy, organ transplant or immunosuppressive therapy)?

Yes ☐ No ☐ Unknown ☐

If Yes, please specify .....

.....

12. Name and address of notifying doctor:.....

.....

.....

.....

Telephone No. .....

13. Name and address of consultant:

.....

.....

.....

.....

Telephone No. .....

14. Other Pertinent Information

a. Lowest total WBC count .....

b. % lymphocytes corresponding to a. .....

c. Other immunological test results (use reverse side of sheet if necessary) .....

.....

.....

d. % of T-helpers among lymphocytes; T-helper/T-suppressor ratio:-

.....

Is patient an IV drug user? Yes ☐ No ☐ Unknown ☐

If Yes, please specify (e.g. heroin) .....

16. a) Did patient have a blood transfusion before this illness?

Yes ☐ No ☐

b) DID PATIENT DONATE BLOOD IN 5 YEARS PRIOR TO ONSET OF THIS ILLNESS?

Yes ☐ No ☐

17. Is this case related to any other known cases?

Yes ☐ No ☐ Unknown ☐

If Yes, how? .....

18. Before the onset of illness, did patient ever have:-

	Yes	No	Unknown	How many times	Year of last infection
a. Gonorrhoea	☐	☐	☐	.....	.....
b. Syphilis	☐	☐	☐	.....	.....
c. Non-gonococcal urethritis	☐	☐	☐	.....	.....
d. Genital herpes simplex	☐	☐	☐	.....	.....
e. Amoebiasis	☐	☐	☐	.....	.....
f. Giardiasis	☐	☐	☐	.....	.....
g. Hepatitis A (infectious)	☐	☐	☐	.....	.....
h. Hepatitis B (serum)	☐	☐	☐	.....	.....
i. Vaccination against hepatitis B	☐	☐	☐	.....	.....

19. Does the patient ever use:-

	Yes	No	Unknown
a. Marijuana	☐	☐	☐
b. Cocaine	☐	☐	☐
c. Heroin	☐	☐	☐
d. Amphetamines	☐	☐	☐
e. Barbiturates	☐	☐	☐

	Yes	No	Unknown
LSD (acid)	☐	☐	☐
Inhalant sexual stimulants (poppers - amyl or butyl nitrite)	☐	☐	☐
Ethyl chloride	☐	☐	☐

20. Sexual Contact with Americans

a. Did the patient travel to the USA or Caribbean countries in the two years prior to illness?

Yes ☐ No ☐ Unknown ☐

If Yes, to what cities and when?

	Month/Year
.....	.....
.....	.....
.....	.....
.....	.....

b. Did the patient have any sexual contacts while travelling?

Yes ☐ No ☐ Unknown ☐

If Yes, please specify:-

City	Month/Year	No. of Contacts
.....	.....	.....
.....	.....	.....
.....	.....	.....
.....	.....	.....

c. Did the patient have any sexual contacts with Americans or Caribbean nationals?

Yes ☐ No ☐ Uknown ☐

City	Month/Year	No. of Contacts
.....	.....	.....
.....	.....	.....
.....	.....	.....
.....	.....	.....

d. Did any regular sex partner of the patient travel to the USA or Caribbean countries, or have sex with Americans or Caribbean nationals outside the USA in the two years prior to illness?

Yes ☐ No ☐ Unknown ☐

If Yes, please specify:-

City	Month/Year
.....	.....
.....	.....
.....	.....

21. Do you know of other patients who might have Kaposi's Sarcoma or any of the other infections mentioned?

Yes ☐ No ☐

If Yes, please specify:-

Number	City/Country
.....	.....
.....	.....
.....	.....

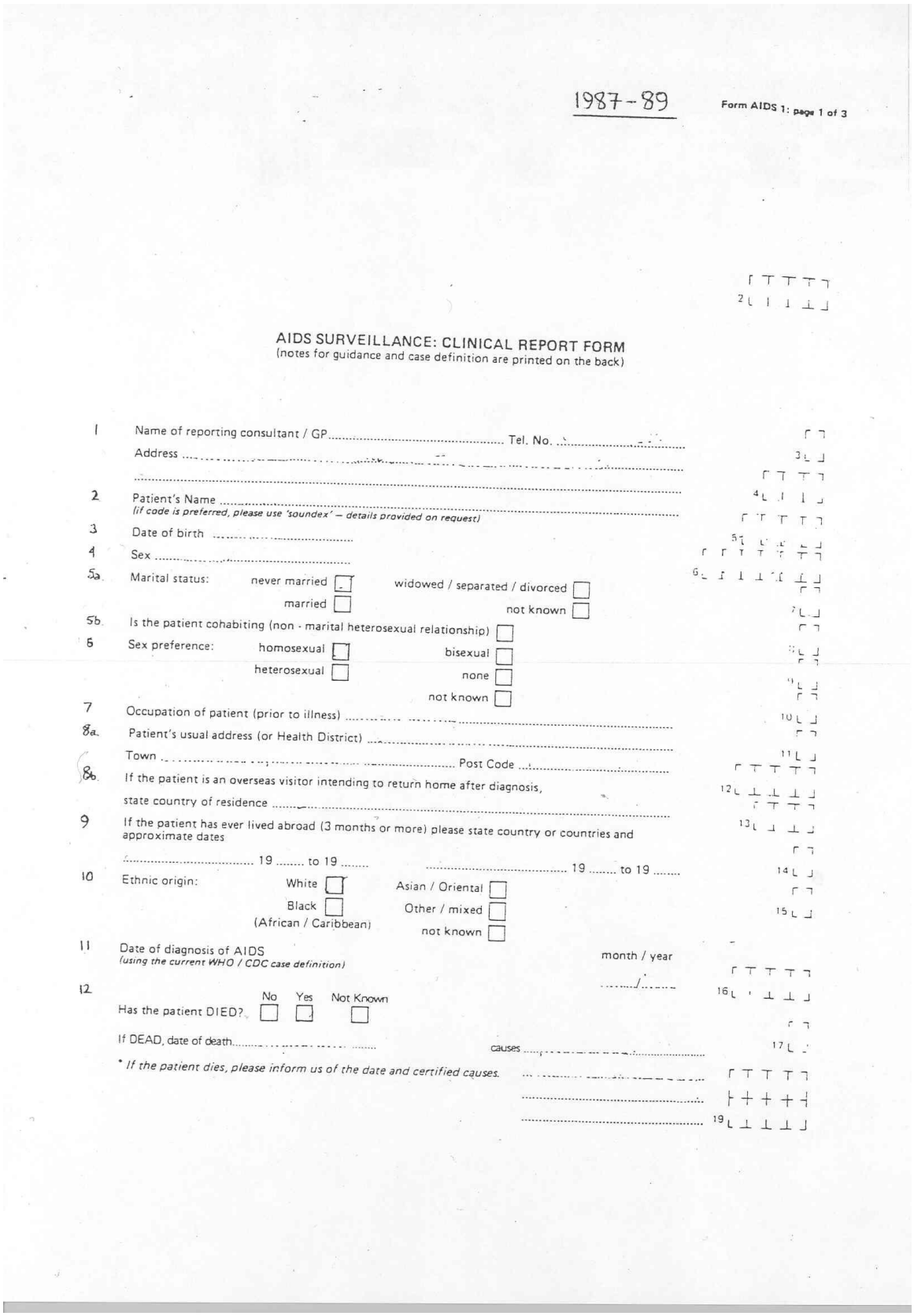
Thank you for your help.

Please return completed questionnaires to:

Dr. M. B. McEvoy
CDSC
61 Colindale Avenue
London NW9 5EQ

Date .....

Form B.2c. ‘AIDS form 87-89’ – ‘AIDS surveillance: clinical report’



1987-89

Form AIDS 1: page 1 of 3

AIDS SURVEILLANCE: CLINICAL REPORT FORM

(notes for guidance and case definition are printed on the back)

1 Name of reporting consultant / GP ..... Tel. No. .....

Address .....

.....

2 Patient's Name .....

(if code is preferred, please use 'soundex' – details provided on request)

3 Date of birth .....

4 Sex .....

5a Marital status: never married ☐ widowed / separated / divorced ☐

married ☐ not known ☐

5b Is the patient cohabiting (non - marital heterosexual relationship) ☐

6 Sex preference: homosexual ☐ bisexual ☐

heterosexual ☐ none ☐

not known ☐

7 Occupation of patient (prior to illness) .....

8a Patient's usual address (or Health District) .....

Town ..... Post Code .....

8b If the patient is an overseas visitor intending to return home after diagnosis, state country of residence .....

9 If the patient has ever lived abroad (3 months or more) please state country or countries and approximate dates

..... 19 ..... to 19 .... 19 ..... to 19 .....

10 Ethnic origin: White ☐ Asian / Oriental ☐

Black (African / Caribbean) ☐ Other / mixed ☐

not known ☐

11 Date of diagnosis of AIDS *(using the current WHO / CDC case definition)* month / year ...../.....

12 Has the patient DIED? No ☐ Yes ☐ Not Known ☐

If DEAD, date of death..... causes .....

* *If the patient dies, please inform us of the date and certified causes.*

POSSIBLE MODE OF INFECTION

PLEASE TICK APPROPRIATE COLUMNS: N = NO, Y = YES, NK = NOT KNOWN

BEFORE ONSET OF ILLNESS — *HAS THE PATIENT EVER*	N	Y	NK	Approx. dates	In Britain	Abroad state Country	
13. Misused drugs by injection							20
14. Had any injections abroad (specify)							22
15. Had blood transfusion							23
16. Had blood products a) for coagulation disorders (specify) ..... b) immunoglobulins etc. (specify)							25
17. Had surgical procedures (specify)							27
18. Been sexually active abroad: state homo / hetero							29
19. Had other known exposure to an HIV+ve person (state type of contact)							31
20. Other possible exposure (eg. ear-piercing, acupuncture needle injury, tattoo, etc.)							33

21. Please give any other comments about exposure

35
36
37
38
39

Please describe clinical features on next page.

22. HIV +ve. Has the patient been found antibody +ve / culture/antigen +ve

	N	Y	NK	NT
antibody +ve				
culture/antigen +ve				

If known: date of last –ve test ...../.....

date of first +ve test ...../.....

(N = NO, Y = YES, NK = NOT KNOWN, NT = NOT TESTED)

23. Any conditions which could lead to immune deficiency other than AIDS/HIV infection

.....

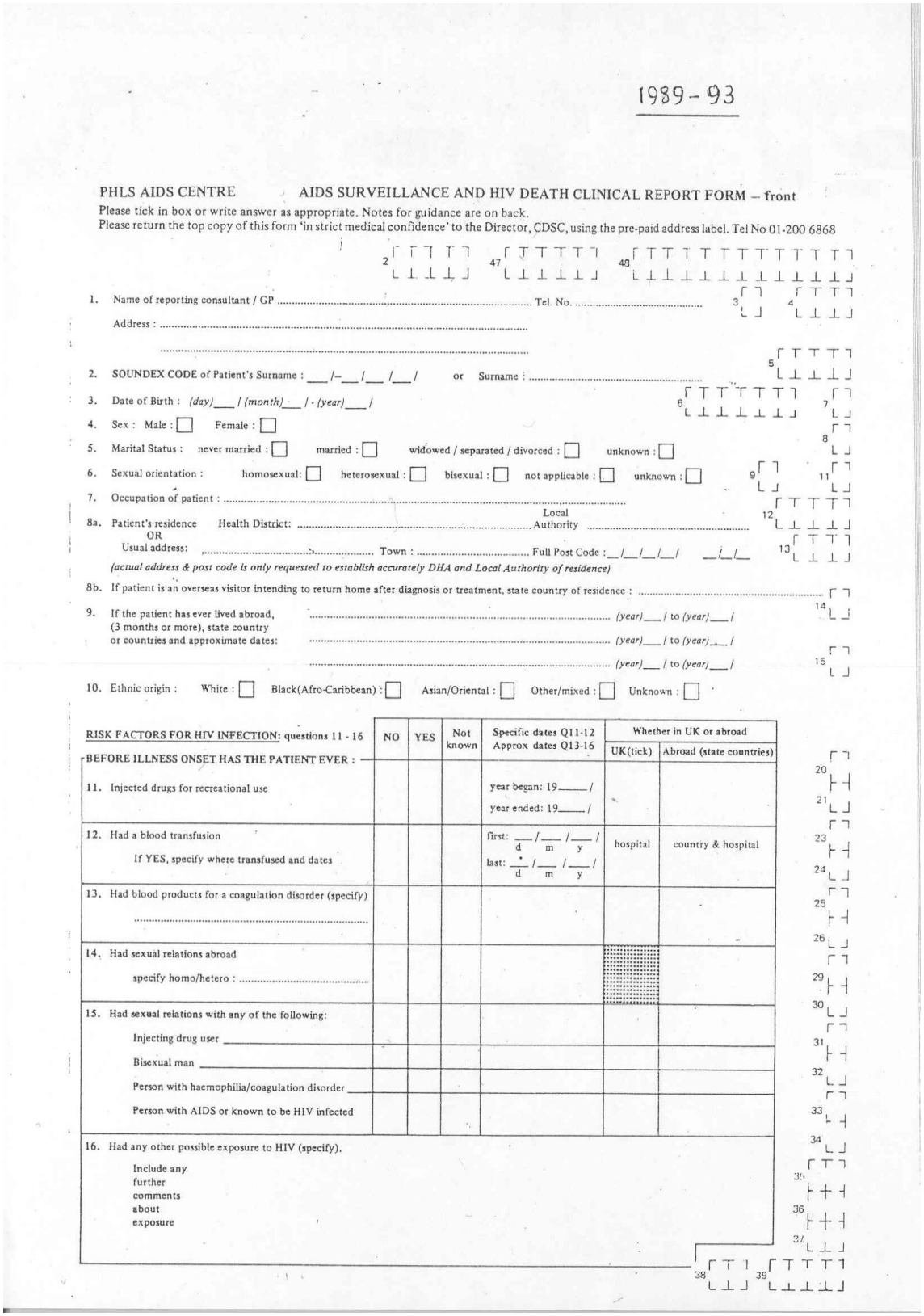
24. CLINICAL FEATURES (CURRENT AND PREVIOUS)

	DISEASE & SITE & ORGANISM (if applicable)	Onset mth/yr	Diagnostic method:– (e.g. culture, serology, biopsy, imaging)
AIDS diagnostic or AIDS – related (opportunistic disorders, skin disease, tumours, T4 cell deficiency, weight loss, lymphadenopathy etc.)	Kaposi's sarcoma		
	Pneumocystis carinii pneumonia		
	CNS involvement (specify)		
	Lymphoma		
	other (specify)		
ANY OTHER Medical conditions requiring out-patient or in-patient treatment (skin rashes, tumours, infections, etc.)	Hepatitis B		
	Gonorrhoea		
	Syphilis		
	Herpes		
	NSGI		
	TB		
	other (specify)		

Form completed by.....

Date.....

Form B.2d. ‘AIDS form 89-93’ – ‘AIDS surveillance and HIV death clinical report’



1989-93

PHLS AIDS CENTRE AIDS SURVEILLANCE AND HIV DEATH CLINICAL REPORT FORM – front

Please tick in box or write answer as appropriate. Notes for guidance are on back.
Please return the top copy of this form ‘in strict medical confidence’ to the Director, CDSC, using the pre-paid address label. Tel No 01-200 6868

1. Name of reporting consultant / GP ..... Tel. No. .....
Address : .....
.....

2. SOUNDEX CODE of Patient's Surname : ___/___/___/___/ or Surname : .....

3. Date of Birth : *(day)*___ / *(month)*___ / - *(year)*___ /

4. Sex : Male : ☐ Female : ☐

5. Marital Status : never married : ☐ married : ☐ widowed / separated / divorced : ☐ unknown : ☐

6. Sexual orientation : homosexual: ☐ heterosexual : ☐ bisexual : ☐ not applicable : ☐ unknown : ☐

7. Occupation of patient : .....

8a. Patient's residence Health District: ..... Local Authority .....
OR
Usual address: ..... Town : ..... Full Post Code : __/__/__/__/ __/__/__
(actual address & post code is only requested to establish accurately DHA and Local Authority of residence)

8b. If patient is an overseas visitor intending to return home after diagnosis or treatment, state country of residence : .....

9. If the patient has ever lived abroad, (3 months or more), state country or countries and approximate dates: ..... *(year)*___ / to *(year)*___ /
..... *(year)*___ / to *(year)*___ /
..... *(year)*___ / to *(year)*___ /

10. Ethnic origin : White : ☐ Black(Afro-Caribbean) : ☐ Asian/Oriental : ☐ Other/mixed : ☐ Unknown : ☐

RISK FACTORS FOR HIV INFECTION: questions 11 - 16 BEFORE ILLNESS ONSET HAS THE PATIENT EVER :	NO	YES	Not known	Specific dates Q11-12 Approx dates Q13-16	Whether in UK or abroad UK(tick)	 Abroad (state countries)
11. Injected drugs for recreational use				year began: 19____ / year ended: 19____ /		
12. Had a blood transfusion If YES, specify where transfused and dates				first: __ / __ / __ / (d m y) last: __ / __ / __ / (d m y)	hospital	country & hospital
13. Had blood products for a coagulation disorder (specify)						
14. Had sexual relations abroad specify homo/hetero :						
15. Had sexual relations with any of the following: Injecting drug user Bisexual man Person with haemophilia/coagulation disorder Person with AIDS or known to be HIV infected						
16. Had any other possible exposure to HIV (specify). Include any further comments about exposure						

Use to report: either: i) Newly diagnosed AIDS cases or ii) Death in an HIV- infected person without AIDS

17a. Laboratory HIV test results:

	Positive	Negative	Inconclusive	Not done
HIV serum antibody test	☐	☐	☐	☐
HIV serum antigen test	☐	☐	☐	☐
Other HIV test? specify:	☐	☐	☐	☐

40 ⌐ ¬ ⌙ ⌋

41 ⌐ | | | | ⌙ | | | | |

42 ⌐ ¬ ⌙ ⌋

17b. If known; Date last negative test: (month)___/(year)___/ Date first positive test: (month)___/(year)___/

18. When AIDS was diagnosed, what was the absolute T-helper(CD4) lymphocyte count _____ $x10^9/l$ Not available: ☐

CLINICAL FEATURES: questions 19 - 24

19. Did the patient have any condition(s), other than the HIV infection, which could cause immune deficiency No: ☐ Yes: ☐ Unknown: ☐

If YES, specify: .....

20. Was an AIDS indicator disease diagnosed either definitively or presumptively (see full list on back) No: ☐ Yes: ☐ Unknown: ☐ ⌐ ¬ ¬ ¬ 16 ⌙ ⌋ ⌋ ⌋ ⌋

If YES, a) For the AIDS indicator disease which was diagnosed first, specify the date of diagnosis: (month) ___ /(year)___ /

b) Specify all AIDS indicator diseases which were diagnosed in this patient, mark whether diagnosed definitively or presumptively, and indicate the onset time of symptoms and/or signs specific to each disease.

AIDS INDICATOR DISEASE (NA–not applicable)	DIAGNOSIS Definitive	DIAGNOSIS Presumptive	Onset Month/year
Candidiasis: trachea, bronchi or lungs	☐	NA	__/__/
Candidiasis: oesophageal	☐	☐	__/__/
Cryptococcosis: extrapulmonary	☐	NA	__/__/
Site			
Cryptosporidiosis, with diarrhoea for over 1 month	☐	NA	__/__/
Cytomegalovirus disease (onset after age 1 month) not in liver, spleen or nodes	☐	NA	__/__/
Site			
Cytomegalovirus retinitis with loss of vision	☐	☐	__/__/
HIV encephalopathy	☐	NA	__/__/
Herpes simplex: ulcer(s) for over 1 month or bronchitis, pneumonitis, oesophagitis (onset after age 1 month)	☐	NA	__/__/
Site			
Kaposi's sarcoma	☐	☐	__/__/
Site			
Lymphoid interstit. pneum. or pulmon. lymphoid hyperplasia in child (≤12y)	☐	☐	__/__/
Lymphoma, Burkitt's, or equiv. term	☐	NA	__/__/
Site			
Lymphoma, immunoblastic or equiv.	☐	NA	__/__/
Site			
Lymphoma, primary in brain	☐	NA	__/__/
Mycobacterium avium complex or *M. kansasii* disseminated	☐	☐	__/__/
Site			
M. tuberculosis, extrapulmonary	☐	☐	__/__/
Site			
Mycobacterium of other unidentified species, disseminated	☐	☐	__/__/
Site			
Pneumocystis carinii pneumonia	☐	☐	__/__/
Toxoplasmosis of brain, onset after age 1 month	☐	☐	__/__/
Wasting syndrome due to HIV	☐	NA	__/__/

43 ⌐ ¬ ⌙ ⌋

44 ⌐ ¬ ⌙ ⌋

46 ⌐ ¬ ⌙ ⌋

Other AIDS Indicator Disease(s): .....

(see list on back) - for each: specify disease, site(s), diagnostic method(s), onset date(s) .....

.....

⌐ ¬ ¬ ¬ | | | ⌙ ⌋ ⌋ ⌋

21. If the patient had any other recognised manifestations of HIV infection (see list on back) ..... give details: specify condition, site(s), diagnostic methods onset date(s) .....

.....

.....

⌐ ¬ ¬ ¬ ¬ 45 ⌙ ⌋ ⌋ ⌋ ⌋

22. If the patient had any other unusual or interesting clinical features give details: .....

.....

specify condition, site(s), diagnostic methods & onset date(s) (include post-mortem findings, cause(s) of death etc., if appropriate) .....

.....

23. Were any of the following conditions ever diagnosed in this patient ⌐ ¬ 17 ⌙ ⌋

	No	Yes	Mth/yr first diagnosed		No	Yes	Mth/yr first diagnosed
Herpes zoster	☐	☐	__/__/	Syphilis	☐	☐	__/__/
Pulmonary TB	☐	☐	__/__/	Gonorrhoea	☐	☐	__/__/
Acute Hepatitis B	☐	☐	__/__/	NSGI	☐	☐	__/__/

18 ⌐ ¬ ¬ ¬ ¬ ⌙ ⌋ ⌋ ⌋ ⌋ ⌋

19 ⌐ ¬ ¬ ¬ ¬ ⌙ ⌋ ⌋ ⌋ ⌋ ⌋

24. Has the patient DIED No: ☐ Yes: ☐ unknown: ☐ If DEAD, specify date of death: (day) ___ /(month)___ /(year)___ /

If ALIVE, and death occurs subsequently, please inform CDSC of the date of death and the certified causes

Form completed by: ..... Position: ..... Date: .....

PHLS AIDS CENTRE

AIDS SURVEILLANCE AND HIV DEATH CLINICAL REPORT FORM – back

What to report on this form

i). All newly diagnosed AIDS cases defined by the presence of an indicator disease (see list below), whether alive or dead. If a patient dies after a report form was completed, please forward a short letter specifying the patient's soundex code (or surname), sex, full dates of birth and death and the certified causes of death.

OR

ii). Death in an HIV infected person in whom no AIDS indicator disease (see list below) was diagnosed either definitively or presumptively.

Information on HIV disease in children is collected via the British Paediatric Surveillance Unit. Paediatricians should make initial reports using the BPSU monthly report card.

To whom should report be sent

i). Return the front blue copy of the AIDS clinical report form 'in strict medical confidence' to:

The Director, PHLS Communicable Disease Surveillance Centre,
61 Colindale Avenue, London NW9 5EQ. ..... Use the pre-paid address label supplied

ii). The middle pink copy may be sent 'in strict medical confidence' to:

The Physician coordinating AIDS/HIV information in the Health District or Region - usually the consultant responsible for communicable disease control (see HC[88]64)

iii). The bottom white copy may be kept by the reporting clinician as a confidential record.

Alternatively, either copies ii or iii, or both, should be destroyed.

Q2 - Soundex surname code

Reporting doctors are strongly encouraged to use the 'soundex code' as an alternative to supplying the patient's surname. The code comprises a set of twelve rules whereby a surname is encoded, either by hand or using a programmable calculator, to a first letter followed by three digits. As no surname is unique to a particular code, confidentiality is assured. When the code is used together with date of birth and sex, however, probable duplicate reports can be readily detected and further enquiries made. For further details contact either CDSC or the physician coordinating AIDS/HIV information in the Health District.

Q3 - Dates

For date of birth and date of death please give day/month/year. Elsewhere month/year or approximate date is requested.

Q8 - Place of residence

District Health Authority and Local Authority of residence is required. In most small towns the name of the town or the first part of the post-code is sufficient to identify this, but in large towns (particularly in London) a full post code is needed because the first part of the post-code may span two or more boundaries.

Q11 to 16 - Risk factors for HIV infection

Reliable information about a patient's risk factors for HIV infection is especially valuable. Where risk factor information is not recorded in the patient's clinical notes, the AIDS clinical report form should be completed by a physician with personal knowledge of the patient or the patient should be reinterviewed.

Q20a - Diagnosis date

For the AIDS indicator disease which was diagnosed first, give the month and year when the result of the definitive diagnostic method became available or when the criteria for a presumptive diagnosis were met.

Q20b - AIDS indicator diseases

Report all AIDS indicator diseases (see list) which have been diagnosed either definitively or presumptively at the time of reporting.

Q20b - Onset times

Give onset time (month and year) of symptoms and/or signs specific to each condition.

Q20b - Definitive Diagnostic Methods & Presumptive Diagnostic Criteria for AIDS Indicator Diseases

Unless otherwise stated, ticking a box for a particular indicator disease overleaf confirms that the disease was diagnosed by a definitive diagnostic method or that all the appropriate criteria for a presumptive diagnosis were met.

AIDS INDICATOR DISEASE	Definitive or Presumptive	Definitive diagnostic method or presumptive diagnostic criteria
Bacterial infections, multiple or recurrent in a child aged 12 years or less*	Definitive	culture, antigen detection, CSF microscopy.
Candidiasis, trachea, bronchi or lungs	Definitive	gross inspection at endoscopy/post-mortem or by microscopy (histology or cytology).
Candidiasis of oesophagus	Definitive	gross inspection at endoscopy/ post-mortem or by microscopy (histology or cytology).
	Presumptive	recent onset retrosternal pain on swallowing or radiological evidence and confirmed oral or pharyngeal candidiasis.
Coccidioidomycosis, disseminated or extrapulmonary	Definitive	microscopy, culture of, or antigen detection in, affected tissue.
Cryptococcosis, extrapulmonary	Definitive	microscopy, culture of, or antigen detection in, affected tissue.
Cryptosporidiosis, with diarrhoea for over 1 month.	Definitive	stool microscopy.
Cytomegalovirus disease (onset after age 1 month) not in liver, spleen or nodes	Definitive	culture lung tissue, microscopy (histology or cytology), antigen or nucleic acid detection.
Cytomegalovirus retinitis	Presumptive	loss of vision and characteristic appearance on serial opthalmoscopy, progressing over several months.
HIV encephalopathy (dementia)*	Definitive	disabling cognitive and/or motor dysfunction, or milestone loss in a child, and no other causes by CSF exam and brain imaging or post-mortem.

Use to report: either: i) Newly diagnosed AIDS cases or ii) Death in an HIV-infected person without AIDS

AIDS INDICATOR DISEASE (continued)	Definitive or Presumptive	Definitive diagnostic method or presumptive diagnostic criteria.
Herpes simplex: ulcers for 1 month or bronchitis, pneumonitis, oesophagitis (onset after age 1 month)	Definitive	culture, microscopy of, or antigen detection in, affected tissue.
Histoplasmosis, disseminated or extrapulmonary	Definitive	microscopy, culture of, or antigen detection in, affected tissue.
Isosporiasis, with diarrhoea for over 1 month	Definitive	microscopy (histology or cytology).
Kaposi's sarcoma	Definitive Presumptive	microscopy (histology or cytology). characteristic erythematous/violaceous plaque-like lesion on skin or mucous membrane.
Lymphoid interstitial pneumonia and/or pulmonary lymphoid hyperplasia in a child aged 12 years or less*	Definitive Presumptive	microscopy (histology or cytology). diffuse bilateral reticulonodular pulmonary interstitial infiltrates for over 2 months and no pathogen identified and no antibiotic response.
Lymphoma: Burkitt's or immunoblastic or primary in brain*	Definitive	microscopy (histology or cytology).
Mycobacteriosis (including extrapulmonary TB) disseminated	Definitive	culture.
Mycobacteriosis disseminated*	Presumptive	AFB (species not identified by culture) on microscopy of stool specimen or normally sterile body fluid/tissue, not lungs, skin, cervical or hilar nodes.
Pneumocystis carinii pneumonia	Definitive Presumptive	microscopy (histology or cytology). recent onset dyspnoea on exertion or dry cough, and diffuse bilateral interstitial infiltrates on CXR and pO_2 <70mm Hg (9.3kPa) and no evidence of bacterial pneumonia.
Progressive multifocal leukoencephalopathy	Definitive	electron microscopy, antigen detection in brain or urine, antibody in serum or CSF.
Salmonella (non-typhoid) septicaemia, recurrent	Definitive	culture.
Toxoplasmosis of brain onset after age 1 month	Definitive Presumptive	microscopy (histology or cytology), mouse inoculation, tissue culture. recent onset focal neurological abnormality or reduced level of conciousness, and mass effect lesion on scan, and serological evidence or specific therapy response.
Wasting syndrome due to HIV	Definitive	weight loss (over 10% baseline) with no other cause, and 30 days or more of either diarrhoea or weakness with fever.

*Further notes on AIDS indicator neoplasms. Mycobacteriosis as an AIDS indicator disease, and on AIDS indicator diseases in children are available from CDSC.

Q21 - OTHER MANIFESTATIONS OF HIV INFECTION - CDC classification

Group I: Acute HIV infection

Group II: Asymptomatic infection

Group III: Persistent Generalised Lymphadenopathy - palpable lymphadenopathy (1cm or greater) at 2 or more extra inguinal sites, for 3 months or more, and no other cause.

Group IV – not an AIDS indicator disease:

Constitutional symptoms/signs which do not meet the criteria for Wasting syndrome due to HIV, give details.
e.g.: disabling weakness, fever.
Herpes zoster - specify whether single or multi-dermatomal.
Infectious diseases not defined as an AIDS indicator infection, give details.
e.g.: extra-intestinal strongyloidosis, nocardiosis, pulmonary tuberculosis.
Myelopathy - unexplained.
Neoplasms not defined as AIDS indicator neoplasms, give details.
Oral candidiasis.
Oral hairy leukoplakia.
Peripheral neuropathy - unexplained.
Seborrhoeic dermatitis.
Thrombocytopenia - idiopathic.

Any other clinical feature attributable to HIV infection, give details.

Further copies

AIDS clinical report forms, copies of the current case definition (WHO Wkly Epidem Rec 1988; 63: 1-8 or Morbid Mortal Wkly Rep 1987: 36: suppl no IS) and details of the soundex code are obtainable from the PHLS AIDS Centre at CDSC (01-200 6868) or CD(S)U (041-946 7120) or from the physician co-ordinating AIDS/HIV information in the Health District or Region.

Surveillance information output

Analyses of data are made monthly and results are circulated in the Communicable Disease Report from CDSC, reported in the medical press, and some of the tabulations are released to the Press by the Department of Health.

THANK YOU FOR YOUR HELP

Form B.2e. ‘AIDS form 93-96’ – ‘AIDS surveillance and HIV death clinical report’

1993 – JULY 1996

PHLS AIDS CENTRE AIDS SURVEILLANCE AND HIV DEATH CLINICAL REPORT FORM

Use to report either: i) A newly diagnosed AIDS case: or ii) Death in an HIV-infected person without AIDS:

In completing this form please tick in box or write answer as appropriate. See white part of this form or ring CDSC, (081-200-6868 Ext. 4453) for guidance. Please return the top copy of this form ‘in strict medical confidence’ to the Director, CDSC using the pre-paid address label.

1. Name of reporting consultant/GP ..... Tel No: .....

   Address: .....

   .....

2. SOUNDEX CODE of Patient’s Surname: ..... — ...../ ...../ ..... or Surname ..... Initial(s) .....

3. Date of Birth: dd/mm/yy ...../ ...../ .....

4. Sex: Male: ☐ 1 Female: ☐ 2

5. Marital Status: never married: ☐ 1 currently married: ☐ 2 widowed/separated/divorced: ☐ 3 unknown: ☐ 9

6. Sexuality: homosexual: ☐ 1 heterosexual: ☐ 2 bisexual: ☐ 3 not applicable: ☐ 4 unknown: ☐ 9

7. Occupation of patient: .....

8. Residence of patient:
   Patient’s usual address: .....

   Town: ..... Full Postcode: .....

   Health district: ..... Local Authority .....
   (address or full postcode are required to establish DHA & Local Authority of residence.)

9. If patient is an overseas visitor in this country for diagnosis or treatment only, please state country of residence : .....

10. If the patient has ever lived abroad, (3 months or more) state country or countries including country of birth, if known, and approximate dates:

    Country(s): ..... From (year) ..... to (year) .....

    Country(s): ..... From (year) ..... to (year) .....

11. Ethnic group: White ☐ 1 Black-Caribbean ☐ 5 Black-African ☐ 6 Black-Other ☐ 7 Indian/Pakistani/Bangladeshi ☐ 8 Other/mixed ☐ 4

EXPOSURE(S) of patient (tick a box for each of Q12 to 17)		Yes	No	N/K	Dates
12. **Sexual intercourse between men**					
13. **Sexual intercourse between men & women:**	give details below of risk factor(s) of sexual partner(s). If patient exposed abroad give possible country(s) of infection and period(s) of exposure.				
14. **Injecting drug use:**	state below if the person ever injected with a syringe and/or needle previously used by someone else.				
15. **Child of HIV infected mother:**	give mother’s date of birth and likely risk factors below.				
16. **Haemophilia/coagulation disorder:**	**specify**				
17. **Blood/tissue recipient:**	state below country as well as date and reason for transfusion.				

18. **Please include any further details about likely exposure**(s)

    .....

    .....

    .....

    .....

19. **HIV test results:**
    Anti-HIV-1 positive: Yes ☐ 2 No ☐ 7 Not tested ☐ 1 Date of first **positive** test: mm/yy ...../ .....

    Anti HIV-2 infected: Yes ☐ 0 No ☐ Other laboratory evidence of HIV infection: .....

20. **If known:** Date of last **negative** HIV test: mm/yy ...../ .....

21. Please tick all the AIDS indicator diseases that have been diagnosed and give a date of diagnosis.

AIDS INDICATOR DISEASE (NA-not applicable)	DIAGNOSIS Definitive	DIAGNOSIS Presumptive	Diagnosis date Month/year
Bacterial infections (multiple) in a child aged less than 13 years	☐	☐	...../.....
Candidiasis: trachea, bronchi or lungs	☐	NA	...../.....
Candidiasis: oesophageal	☐	☐	...../.....
Cervical carcinoma, invasive	☐	NA	...../.....
Coccidioidomycosis: extrapulmonary	☐	NA	...../.....
Cryptococcosis: extrapulmonary	☐	NA	...../.....
Cryptosporidiosis: with diarrhoea for over 1 month	☐	NA	...../.....
Cytomegalovirus retinitis	☐	☐	...../.....
CMV disease not in liver, spleen or nodes	☐	NA	...../.....
Encephalopathy (dementia) due to HIV	☐	NA	...../.....
Herpes simplex ulcer(s) for over 1 month or bronchitis, pneumonitis, oesophagitis	☐	NA	...../.....
Histoplasmosis: disseminated or extrapulmonary	☐	☐	...../.....
Isosporiasis: with diarrhoea for over 1 month	☐	☐	...../.....
Kaposi's sarcoma	☐	☐	...../.....

AIDS INDICATOR DISEASE (NA-not applicable)	DIAGNOSIS Definitive	DIAGNOSIS Presumptive	Diagnosis date Month/year
Lymphoid interstit. pneum. or pulmon. lymphoid hyperplasia in child (<13y)	☐	☐	...../.....
Lymphoma, Burkitt's, or equiv. term	☐	NA	...../.....
Lymphoma, immunoblastic or equiv.	☐	NA	...../.....
Lymphoma, primary in brain	☐	NA	...../.....
Mycobacterium avium: extrapulmonary	☐	☐	...../.....
M. tuberculosis: pulmonary	☐	☐	...../.....
M. tuberculosis: extrapulmonary	☐	☐	...../.....
Mycobacterium of other or unidentified species, disseminated	☐	☐	...../.....
Pneumocystis carinii pneumonia	☐	☐	...../.....
Pneumonia: recurrent within a twelve month period	☐	☐	...../.....
Prog. multifocal leukoencephalopathy	☐	☐	...../.....
Salmonella septicaemia, recurrent	☐	☐	...../.....
Toxoplasmosis of brain	☐	☐	...../.....
Wasting syndrome due to HIV	☐	NA	...../.....

22. Does the patient have any cause of immunodeficiency other than HIV infection? Yes ☐ No ☐ Not known ☐

If yes, please specify .....

23. Give CD4 lymphocyte count only if taken within 3 months of the AIDS diagnosis date?

CD4 count ..... /µl Date of CD4 count: mm/yy ...../.....

24. If the patient had a clinical manifestation of HIV infection prior to an AIDS indicator disease please specify the initial illness and its diagnosis date (see illnesses in classification system on back of form):

Illness ..... Date: mm/yy ...../.....

25. If the patient has had treatment **for at least 3 months** at some point **within the 2 years** prior to the diagnosis of AIDS, please specify the drugs used below:

Anti retroviral drugs	Drugs for prophylaxis of opportunistic infections
1.	1. .....
2.	2. .....
3.	3. .....

26. Has the patient DIED: Yes ☐ 2 No ☐ 1 Not known ☐ 9

If YES, please specify date of death: dd/mm/yy ...../...../.....

Cause of death (if known): .....

If currently ALIVE, please inform CDSC of date and cause of death should it subsequently occur.

Completed by: ..... **Position:** ..... **Date:** ...../...../.....

AIDS INDICATOR DISEASE*	Definitive or Presumptive	Definitive diagnostic method(s) or presumptive diagnositic criteria.
Bacterial infections, multiple or recurrent, in a child aged less than 13 years	Definitive	culture, antigen detection, CSF microscopy.
Candidiasis, trachea, bronchi or lungs	Definitive	gross inspection at endoscopy/post-mortem or by microscopy (histology or cytology).
Candidiasis of oesophagus	Definitive Presumptive	gross inspection at endoscopy/post-mortem or by microscopy (histology or cytology). recent onset retrosternal pain on swallowing or radiological evidence and confirmed oral or pharyngeal candidiasis.
Cervical carcinoma, invasive	Definitive	histology.
Coccidioidomycosis, disseminated or extrapulmonary	Definitive	microscopy, culture of, or antigen detection in, affected tissue.
Cryptococcosis, extrapulmonary	Definitive	microscopy, culture of, or antigen detection in, affected tissue.
Cryptosporidiosis, with diarrhoea for over 1 month	Definitive	stool microscopy.
Cytomegalovirus retinitis	Presumptive	loss of vision and characteristic appearance on serial ophthalmoscopy, progressing over several months.
Cytomegalovirus disease (onset after age 1 month) not in liver, spleen or nodes	Definitive	culture lung tissue, microscopy (histology or cytology), antigen or nucleic acid detection.
Encephalopathy (dementia) due to HIV	Definitive	HIV infection and disabling cognitive and/or motor dysfunction, or milestone loss in a child, with no other causes by CSF examination, brain imaging or post-mortem.
Herpes simplex: ulcers for 1 month or bronchitis, pneumonitis, oesophagitis (onset after age 1 month)	Definitive	culture, microscopy of, or antigen detection in, affected tissue.
Histoplasmosis, disseminated or extrapulmonary	Definitive	microscopy, culture of, or antigen detection in, affected tissue.
Isosporiasis, with diarrhoea for over 1 month	Definitive	microscopy (histology or cytology).
Kaposi's sarcoma	Definitive Presumptive	microscopy (histology or cytology). characteristic erythematous/violaceous plaque-like lesion on skin or mucous membrane.
Lymphoid interstitial pneumonia and/or pulmonary lymphoid hyperplasia in a child aged less than 13 years	Definitive Presumptive	microscopy (histology or cytology). diffuse bilateral reticulonodular pulmonary interstitial infiltrates for over 2 months and no pathogen identified and no antibiotic response.
Lymphoma: Burkitt's or immunoblastic or primary in brain*	Definitive	microscopy (histology or cytology).
Mycobacteriosis disseminated* (including extrapulmonary TB)	Definitive Presumptive	culture. AFB (species not identified by culture) on microscopy of stool specimen or normally sterile body fluid/tissue, not lungs, skin, cervical or hilar nodes.
Mycobacteriosis: pulmonary tuberculosis*	Definitive Presumptive	culture or other definitive demonstration of *M. tuberculosis* infection. clinical diagnosis, with or without AFB on microscopy, resulting in initiation of anti-TB therapy.
Pneumocystis carinii pneumonia	Definitive Presumptive	microscopy (histology or cytology). recent onset dyspnoea on exertion or dry cough, and diffuse bilateral interstitial infiltrates on CXR and pO2 <70mm Hg (9.3kPa) and no evidence of bacterial pneumonia.
Pneumonia recurrent within a 12 mth period	Definitive Presumptive	two episodes proven microbiologically. CXR or clinical diagnoses of two distinct episodes of pneumonia.
Progressive multifocal leukoencephalopathy	Definitive	electron microscopy, antigen detection in brain or urine, antibody in serum or CSF.
Salmonella (non-typhoid) septicaemia, recurrent	Definitive	culture.
Toxoplasmosis of brain onset after age 1 month	Definitive Presumptive	microscopy (histology or cytology), mouse inoculation, tissue culture. recent onset focal neurological abnormality or reduced level of consciousness, and mass effect lesion on scan, and serological evidence or specific therapy response.
Wasting syndrome due to HIV	Definitive	weight loss (over 10% baseline) with no other cause, and 30 days or more of either diarrhoea or weakness with fever.

*Full case definition and notes on AIDS indicator diseases for neoplasms, mycobacteriosis and indicator diseases in children are available from CDSC.

1993 CLASSIFICATION SYSTEM

Category A: Acute (primary) HIV infection or Asymptomatic HIV infection or Persistent Generalized Lymphadenopathy.

Category B: Symptomatic with conditions other than those included in categories A or C attributed to HIV infection or which are indicative of a defect in cell mediated immunity.

For example:-

Bacillary angiomatosis œ Candidiasis, oropharyngeal (thrush) œ Candidiasis, vulvovaginal; persistent, frequent, or poorly responsive to therapy œ Cervical dysplasia (moderate or severe)/cervical carcinoma in situ œ Constitutional symptoms, such as fever (38.5 C) or diarrhoea lasting > 1 month œ Hairy leukoplakia, oral œ Herpes zoster (shingles), involving at least two distinct episodes or more than one dermatome œ Idiopathic thrombocytopenic purpura œ Listeriosis œ Pelvic inflammatory disease œ Peripheral neuropathy.

Category C: Clinical conditions listed in the AIDS surveillance case definition presumptively or definitively diagnosed (see above).

WHAT TO REPORT ON THIS FORM

1) **Any case of AIDS in a patient (alive or dead).**

An HIV infected patient is defined as having AIDS if they have one or more of the indicator diseases listed in Q21. The following indicator diseases were added to the list in 1993.

i) Pulmonary tuberculosis
ii) Recurrent pneumonia in a twelve month period
iii) Cervical carcinoma, invasive

The 1993 list of AIDS indicator diseases should be applied to all HIV infected persons alive on or after 1 January 1993, who have not yet been reported as having AIDS.

Most people fulfilling the AIDS case definition will have had a positive HIV antibody test result. In people who have refused an HIV test, definitive methods of diagnosis for AIDS indicator diseases are generally necessary. Where necessary a consultant at the PHLS AIDS Centre at CDSC can be contacted to discuss the case definition.

2) **Any death in an HIV infected person in whom an AIDS indicator disease has not been diagnosed.**

WHAT TO DO WITH THIS FORM

1) Return the blue copy of the report form 'in strict medical confidence' to:

The Director, PHLS Communicable Disease Surveillance Centre,
61 Colindale Avenue, London NW9 5EQ. **Use the pre-paid address label supplied.**

2) The pink copy may be sent 'in strict medical confidence' to:

The Physician coordinating AIDS/HIV information in the Health District or Region
- usually the consultant responsible for communicable disease control (see HC[88]64)

3) The white copy may be kept by the reporting clinician as a confidential record.

If not used as suggested above, copies 2 or 3 should be destroyed.

GUIDANCE FOR COMPLETING THE FORM

Q2 - Soundex code of surname

Reporting doctors are encouraged to use the 'soundex code' as an alternative to giving the patient's surname. The soundex code is derived by following a set of eight rules to convert the surname to a first letter followed by three digits. As no surname is unique to a particular soundex code, confidentiality is assured. When the code is used with date of birth and sex, probable duplicate reports can be readily detected and further enquiries made if necessary. The soundex code can be derived by hand, by using a programmed calculator or with a short computer programme available free of charge from CDSC. For further details of soundex coding contact either CDSC or the physician coordinating AIDS/HIV information in the Health District.

Q3 - Dates

For date of birth and date of death please give day/month/year. Elsewhere month/year or approximate date is requested.

Q8 - Place of residence

District Health Authority and Local Authority of residence are required. In many small towns the name of the town or the first part of the postcode is sufficient to identify this, but in large towns (particularly in London) a full postcode is needed because the first part of the postcode may span administrative boundaries. Postcodes are especially useful if, in future, district boundaries change.

Q12 to 18 - Exposure category

Reliable information about a patient's risk factors for HIV infection is especially important. Where risk factor information is not recorded in the patient's clinical notes, the AIDS clinical report form should be completed by a physician with personal knowledge of the patient. The patient may need to be re-interviewed.

Q21 - AIDS indicator diseases

Tick all the AIDS indicator diseases (see list) which have been diagnosed. For each disease give the month and year when the positive result of the definitive diagnostic method became available or when the criteria for a presumptive diagnosis were met.

Unless otherwise stated, ticking a box for an indicator disease listed confirms that the disease was diagnosed by a definitive diagnostic method or that all the appropriate criteria for a presumptive diagnosis were met (see diagnostic criteria in panel opposite).

Form B.2f. ‘AIDS form 96-97’ – ‘AIDS surveillance and HIV death clinical report’

PHLS AIDS CENTRE AIDS SURVEILLANCE AND HIV DEATH CLINICAL REPORT FORM

Use to report either: i) A newly diagnosed AIDS case: ☐ or ii) Death in an HIV-infected person without AIDS: ☐

Please tick white boxes or write answer as appropriate. See back of this form or ring CDSC, (0181 200 6868 Ext. 4453) for guidance.

PLEASE DO NOT WRITE IN THIS BOX

DATE A ☐ D ☐ P ☐

NO. E D/P ☐

HIV NO.

INITIALS IND. DIS.

1. Name of reporting consultant/GP .....

Address: .....

.....

Tel No: .....

2. SOUNDEX CODE of Patient's Surname: ..... – ...../ ...../ ..... or Surname ..... Initial(s) .....

3. Date of Birth: dd/mm/yy ...../ ...../ ..... Clinic Number (if known) .....

4. Sex: Male ☐ 1 Female ☐ 2

5. Marital Status: never married ☐ 1 currently married ☐ 2 widowed/separated ☐ 3 divorced ☐ 3 unknown ☐ 9

6. Sexuality: homosexual ☐ 1 heterosexual ☐ 2 bisexual ☐ 3 unknown ☐ 9 not applicable ☐ 4

7. Occupation of patient: .....

8. Residence of patient : (address or full postcode is ONLY required to establish Health Authority & Local Authority of residence).

Patient's usual address: .....

Town: ..... Full Postcode:

Health Authority: Local Authority:

9. Is patient an overseas visitor in this country for diagnosis or treatment only? Yes ☐ No ☐

If yes, please state country of residence :.....

10. If the patient has ever lived abroad (3 months or more), state country or countries including country of birth, if known, and approximate dates:

Country(s): ..... From ..... to..... Country(s): ..... From ..... to .....

11. Ethnic group: White ☐ 1 Black-Caribbean ☐ 5 Black-African ☐ 6 Black-Other ☐ 7

Indian / Pakistani / Bangladeshi ☐ 8 Other/mixed ☐ 4 Unknown ☐ 9

EXPOSURE(S) of patient (tick a box for each of Q12 to 17)		Yes	No	N/K	Dates
12. Sexual intercourse between men					
13. Sexual intercourse between men & women:	give details below of risk factor(s) of sexual partner(s) and country(ies) of possible exposure including UK.				
14. Injecting drug use:	state below if the person NEVER injected with a syringe and/or needle previously used by someone else.				
15. Child of HIV infected mother:	give mother's date of birth and likely risk factors below.				
16. Haemophilia/coagulation disorder: specify					
17. Blood/tissue recipient:	state below country as well as date and reason for transfusion.				

18. Please include any further details about likely exposure(s)

.....

.....

.....

HIV INFECTION:

19. **If known:** Date of last **negative** HIV antibody test: mm/yy ...../.....

20. HIV-1 positive: Yes ☐ 2 No ☐ 7 Not tested ☐ 1 Date of first positive test: mm/yy ...../..... where: .....

HIV-2 positive: Yes ☐ 0 No ☐ Not tested ☐

21. Please tick all the AIDS indicator diseases diagnosed (unless HIV death report) and give a date of diagnosis.

AIDS INDICATOR DISEASE NA-not applicable	DIAGNOSIS Definitive	 Presumptive	DATE mm/yy
Bacterial infections (multiple) in a child aged less than 13 years	☐	NA	...../.....
Candidiasis: trachea, bronchi or lungs	☐	NA	...../.....
Candidiasis: oesophageal	☐	☐	...../.....
Cervical carcinoma, invasive**	☐	NA	...../.....
Coccidioidomycosis: extrapulmonary	☐	NA	...../.....
Cryptococcosis: extrapulmonary	☐	NA	...../.....
Cryptosporidiosis: with diarrhoea for over 1 month	☐	NA	...../.....
Cytomegalovirus retinitis	☐	☐	...../.....
CMV disease not in liver, spleen or nodes	☐	NA	...../.....
Encephalopathy (dementia) due to HIV	☐	NA	...../.....
Herpes simplex: ulcer(s) for over 1 month or bronchitis, pneumonitis, oesophagitis	☐	NA	...../.....
Histoplasmosis: disseminated or extrapulmonary	☐	NA	...../.....
Isosporiasis: with diarrhoea for over 1 month	☐	NA	...../.....
Kaposi's sarcoma	☐	☐	...../.....

AIDS INDICATOR DISEASE NA-not applicable	DIAGNOSIS Definitive	 Presumptive	DATE mm/yy
Lymphoid interstit. pneum. or pulmon. lymphoid hyperplasia in child (<13y)	☐	☐	...../.....
Lymphoma, Burkitt's, or equiv. term	☐	NA	...../.....
Lymphoma, immunoblastic or equiv.	☐	NA	...../.....
Lymphoma, primary in brain	☐	NA	...../.....
Mycobacterium avium: extrapulmonary	☐	☐	...../.....
M. tuberculosis: pulmonary**	☐	☐	...../.....
M. tuberculosis: extrapulmonary	☐	☐	...../.....
Mycobacterium of other or unidentified species, disseminated	☐	☐	...../.....
Pneumocystis carinii pneumonia	☐	☐	...../.....
Pneumonia: recurrent within a twelve month period**	☐	☐	...../.....
Prog. multifocal leukoencephalopathy	☐	NA	...../.....
Salmonella septicaemia, recurrent	☐	NA	...../.....
Toxoplasmosis of brain	☐	☐	...../.....
Wasting syndrome due to HIV	☐	NA	...../.....

**Applicable only to patients aged ≥13 years diagnosed with those conditions on or after 1st January 1993.

22. If patient is not known to be HIV infected please state any non-HIV related cause of immunodeficiency

.....

TREATMENT:

23. Was the patient on anti retroviral therapy before the AIDS diagnosis? Yes ☐ No ☐ Not Known ☐
If known, date of first starting anti retroviral treatment: mm/yy ...../.....

24. Did the patient receive prophylaxis for PCP before the AIDS diagnosis? Yes ☐ No ☐ Not Known ☐
If known, date of first starting anti PCP prophylaxis: mm/yy ...../.....

25. Did the patient receive prophylaxis for any other opportunistic infection before the AIDS diagnosis? Yes ☐ No ☐ Not Known ☐

26. Has the patient DIED: Yes ☐ 2 No ☐ 1

If YES, please specify date of death: dd/mm/yy ...../...../.....

Cause of death (if known): .....

If patient subsequently dies, please inform CDSC by phone (0181 200 6868 ext. 4453) or letter of date and cause of death.

Completed by: ..... Position: ..... Date: ...../...../.....

Please return the completed form 'in strict medical confidence' to:

Director, CDSC, 61 Colindale Avenue, London NW9 5EQ. Pre-paid address labels supplied.

Form B.2g. ‘AIDS form 97-98’ – ‘AIDS surveillance and HIV death clinical report’

PHLS AIDS CENTRE

AIDS SURVEILLANCE AND HIV DEATH CLINICAL REPORT FORM

Use to report either: i) A newly diagnosed AIDS case: ☐ or ii) Death in an HIV-infected person without AIDS: ☐

Please tick white boxes or write answer as appropriate. See back of this form or ring CDSC (0181 200 6868 Ext. 4453) for guidance.

PLEASE DO NOT WRITE IN THIS BOX
DATE A9707
NO.
HIV NO.
INITIALS .... IND. DIS. .....

1. Name of reporting consultant/GP .....
   Address: .....
   Tel No: .....
2. **SOUNDEX CODE** of Patient's Surname: ..... — ..../ ...../ ... ... **or** Surname ..... Initial(s) .....
3. Date of Birth: dd/mm/yy ...../ ...../ .... Clinic Number (if known) .....
4. Sex: Male ☐ 1 Female ☐ 2
5. Marital Status: never married ☐ 1 currently married ☐ 2 widowed/separated/divorced ☐ 3 unknown ☐ 9
6. Sexuality: homosexual ☐ 1 heterosexual ☐ 2 bisexual ☐ 3 unknown ☐ 9 not applicable ☐ 4
7. Occupation of patient: ..... ☐
8. Residence of patient: **(address or full postcode are ONLY required to establish Health Authority and Local Authority of residence).**
   Patient's usual address:.....
   Town: ..... Full Postcode:
   Health Authority: Local Authority:
9. Is patient an ***overseas visitor*** in this country for ***diagnosis or treatment only***? Yes ☐ No ☐
   If **yes**, please state country of residence :.....
10. If the patient has ever lived abroad (3 months or more), state country or countries including country of birth, if known, and approximate dates:
    Country(s): ..... From ..... to..... Country(s): ..... From ..... to .....
11. Ethnic group: White ☐ 1 Black-Caribbean ☐ 5 Black-African ☐ 6 Black-Other ☐ 7
    Indian / Pakistani / Bangladeshi ☐ 8 Other/mixed ☐ 4 Unknown ☐ 9

EXPOSURE(S) of patient (tick a box for each of Q12 to 17)		Yes	No	N/K	Dates
12. **Sexual intercourse between men**					
13. **Sexual intercourse between men & women:**	give details below of risk factor(s) of sexual partner(s) and country(s) of possible exposure including UK.				
14. **Injecting drug use:**	state below if the person NEVER injected with a syringe and/or needle previously used by someone else.				
15. **Child of HIV infected mother:**	give mother's date of birth and likely risk factors below.				
16. **Haemophilia/coagulation disorder:**	**specify**				
17. **Blood/tissue recipient:**	state below country as well as date and reason for transfusion.				
18. **Please include any further details about likely exposure(s)**					

HIV INFECTION:

19. **If known:** Date of last **negative** HIV antibody test: mm/yy ..... / .....
20. HIV-1 positive: Yes ☐ 1 No ☐ 5 Not tested ☐ 8 Date of first **positive** test: mm/yy ..... / ..... where: .....
    HIV-2 positive: Yes ☐ 2 No ☐ Not tested ☐

21. Please tick all the AIDS indicator diseases diagnosed (unless HIV death report) and give date of diagnosis.

AIDS INDICATOR DISEASE (NA-not applicable)	DIAGNOSIS Definitive	DIAGNOSIS Presumptive	DATE mm/yy
1. Bacterial infections (multiple) in a child aged less than 13 years	☐	NA	...../.....
2. Candidiasis: trachea, bronchi or lungs	☐	NA	...../.....
3. Candidiasis: oesophageal	☐	☐	...../.....
4. Cervical carcinoma, invasive**	☐	NA	...../.....
5. Coccidioidomycosis: extrapulmonary	☐	NA	...../.....
6. Cryptococcosis: extrapulmonary	☐	NA	...../.....
7. Cryptosporidiosis: with diarrhoea for over 1 month	☐	NA	...../.....
8. Cytomegalovirus retinitis	☐	☐	...../.....
9. CMV disease not in liver, spleen or nodes	☐	NA	...../.....
10. Encephalopathy (dementia) due to HIV	☐	NA	...../.....
11. Herpes simplex: ulcer(s) for over 1 month or bronchitis, pneumonitis, oesophagitis	☐	NA	...../.....
12. Histoplasmosis: disseminated or extrapulmonary	☐	NA	...../.....
13. Isosporiasis: with diarrhoea for over 1 month	☐	NA	...../.....
14. Kaposi's sarcoma	☐	☐	...../.....
15. Lymphoid interstit. pneum. or pulmon. lymphoid hyperplasia in child (<13y)	☐	☐	...../.....
16. Lymphoma, Burkitt's, or equiv. term	☐	NA	...../.....
17. Lymphoma, immunoblastic or equiv.	☐	NA	...../.....
18. Lymphoma, primary in brain	☐	NA	...../.....
19. *Mycobacterium avium:* extrapulmonary	☐	☐	...../.....
20. *M. tuberculosis*: pulmonary**	☐	☐	...../.....
21. *M. tuberculosis:* extrapulmonary	☐	☐	...../.....
22. *Mycobacterium* of other or unidentified species, disseminated	☐	☐	...../.....
23. *Pneumocystis carinii* pneumonia	☐	☐	...../.....
24. Pneumonia: recurrent within a twelve month period**	☐	☐	...../.....
25. Prog. multifocal leukoencephalopathy	☐	NA	...../.....
26. Salmonella septicaemia, recurrent	☐	NA	...../.....
27. Toxoplasmosis of brain	☐	☐	...../.....
28. Wasting syndrome due to HIV	☐	NA	...../.....

**Applicable only to patients aged ≥13 years diagnosed with those conditions on or after 1st January 1993.

22. **Please state any non-HIV related cause of immunodeficiency**

.....

TREATMENT:

23. Was the patient *ever* on anti retroviral therapy *before* the AIDS diagnosis? Yes ☐ 1 No ☐ 0 Not Known ☐

If "Yes", what was the most intensive level of pre AIDS therapy?

triple or more ☐ 5 dual ☐ 6 mono ☐ 7 Not Known ☐ 4

Were they *still* on anti retroviral therapy *at the time* of the AIDS Diagnosis? Yes ☐ 1 No ☐ 0 Not Known ☐

If "No" please state reason, if known: .....

24. Did the patient receive prophylaxis for PCP *before* the AIDS diagnosis? Yes ☐ 1 No ☐ 0 Not Known ☐

25. Did the patient receive prophylaxis for any other opportunistic infection *before* the AIDS diagnosis? Yes ☐ 1 No ☐ 0 Not Known ☐

26. Has the patient DIED: Yes ☐ 2 No ☐ 1 If NO, please specify the date of last contact dd/mm/yy ...../...../.....

If YES, please specify date of death: dd/mm/yy ..../...../.....

Cause of death (if known): .....

If patient subsequently dies, please inform CDSC by phone (0181 200 6868 ext 4453) or letter of date and cause of death.

Completed by: ..... **Position:** ..... **Date:** ...../...../.....

Please return the completed form 'in strict medical confidence' to:

Director, CDSC, 61 Colindale Avenue, London NW9 5EQ. Pre-paid address labels supplied.

Form B.2h. ‘AIDS form 98-99’ – ‘AIDS surveillance and HIV death clinical report’

PHLS AIDS CENTRE **AIDS SURVEILLANCE AND HIV DEATH CLINICAL REPORT FORM**

Use to report either: i) A newly diagnosed AIDS case: ☐ 2 or ii) Death in an HIV-infected person **without AIDS:** ☐ 4

Please tick white boxes or write answer as appropriate. See back of this form or ring CDSC (0181 200 6868 ext4453) for guidance.

PLEASE DO NOT WRITE IN THIS BOX
DATE A9808
NO.
IND. DIS. ..... DATE .....
INITIALS .....

1. Name of reporting consultant/GP
   Address:
   Tel No:
2. **SOUNDEX CODE** of Patient's Surname: Surname Initial(s)
3. Date of Birth: dd/mm/yy / / Clinic Number (if known)
4. Sex: Male ☐ 1 Female ☐ 2
5. Marital Status: never married ☐ 1 currently married ☐ 2 widowed/separated/divorced ☐ 3 unknown ☐ 9
6. Sexuality: homosexual ☐ 1 heterosexual ☐ 2 bisexual ☐ 3 unknown ☐ 9 not applicable ☐ 4
7. Occupation of patient:
8. Residence of patient: **(address or full postcode are ONLY required to establish Health Authority and Local Authority of residence).**
   Patient's usual address:
   Town: Full Postcode:
   Health Authority: Local Authority:
9. Is patient an ***overseas visitor*** in this country for ***diagnosis or treatment only***? Yes ☐ No ☐
   If **yes**, please state country of residence :
10. If the patient has ever lived abroad (3 months or more), state country or countries including country of birth, if known, and approximate dates:
    Country(s): From to Country(s): From to
11. Ethnic group: White ☐ 1 Black-Caribbean ☐ 5 Black-African ☐ 6 Black-Other ☐ 7
    Indian / Pakistani / Bangladeshi ☐ 8 Other/mixed ☐ 4 Unknown ☐ 9

EXPOSURE(S) of patient (tick a box for each of Q12 to 17)		**Yes**	**No**	**N/K**	**Dates**
12. **Sexual intercourse between men**					
13. **Sexual intercourse between men & women:**	give details below of risk factor(s) of sexual partner(s) and country(s) of possible exposure including UK.				
14. **Injecting drug use:**	state below if the person NEVER injected with a syringe and/or needle previously used by someone else.				
15. **Child of HIV infected mother:**	give mother's date of birth and likely risk factors below.				
16. **Haemophilia/coagulation disorder: specify**					
17. **Blood/tissue recipient**:	state below country as well as date and reason for transfusion.				
18. **Please include any further details about likely exposure(s)**					

HIV INFECTION:

19. **If known:** Date of last **negative** HIV antibody test: mm/yy /
20. HIV-1 positive: Yes ☐ 1 No ☐ 5 Not tested ☐ 8 Date of first **positive** test: mm/yy / where:
    HIV-2 positive: Yes ☐ 2 No ☐ Not tested ☐

21. Please tick any AIDS indicator diseases diagnosed (unless HIV death report) and give date of diagnosis.

AIDS INDICATOR DISEASE NA-not applicable	DIAGNOSIS Definitive	 Presumptive	DATE mm/yy
1. Bacterial infections (multiple) in a child aged less than 13 years	☐	NA	...../.....
2. Candidiasis: trachea, bronchi or lungs	☐	NA	...../.....
3. Candidiasis: oesophageal	☐	☐	...../.....
4. Cervical carcinoma, invasive**	☐	NA	...../.....
5. Coccidioidomycosis: extrapulmonary	☐	NA	...../.....
6. Cryptococcosis: extrapulmonary	☐	NA	...../.....
7. Cryptosporidiosis: with diarrhoea for over 1 month	☐	NA	...../.....
8. Cytomegalovirus retinitis	☐	☐	...../.....
9. CMV disease not in liver, spleen or nodes	☐	NA	...../.....
10. Encephalopathy (dementia) due to HIV	☐	NA	...../.....
11. Herpes simplex: ulcer(s) for over 1 month or bronchitis, pneumonitis, oesophagitis	☐	NA	...../.....
12. Histoplasmosis: disseminated or extrapulmonary	☐	NA	...../.....
13. Isosporiasis: with diarrhoea for over 1 month	☐	NA	...../.....
14. Kaposi's sarcoma	☐	☐	...../.....

AIDS INDICATOR DISEASE NA-not applicable	DIAGNOSIS Definitive	 Presumptive	DATE mm/yy
15. Lymphoid interstit. pneum. or pulmon. lymphoid hyperplasia in child (<13y)	☐	☐	...../.....
16. Lymphoma, Burkitt's, or equiv. term	☐	NA	...../.....
17. Lymphoma, immunoblastic or equiv.	☐	NA	...../.....
18. Lymphoma, primary in brain	☐	NA	...../.....
19. *Mycobacterium avium:* extrapulmonary	☐	☐	...../.....
20. *M. tuberculosis*: pulmonary**	☐	☐	...../.....
21. *M. tuberculosis:* extrapulmonary	☐	☐	...../.....
22. *Mycobacterium* of other or unidentified species, disseminated	☐	☐	...../.....
23. *Pneumocystis carinii* pneumonia	☐	☐	...../.....
24. Pneumonia: recurrent within a twelve month period**	☐	☐	...../.....
25. Prog. multifocal leukoencephalopathy	☐	NA	...../.....
26. Salmonella septicaemia, recurrent	☐	NA	...../.....
27. Toxoplasmosis of brain	☐	☐	...../.....
28. Wasting syndrome due to HIV	☐	NA	...../.....

**Applicable only to patients aged ≥13 years diagnosed with those conditions on or after 1st January 1993.

22. **Please state any non-HIV related cause of immunodeficiency**

TREATMENT: Please show extent of treatment received EITHER before AIDS diagnosis OR before death without AIDS.
Please do NOT report any post-AIDS diagnosis treatment

23. Was the patient ***ever*** on anti retroviral therapy ***before*** the AIDS diagnosis/death without AIDS? Yes ☐ 1 No ☐ 0 N K ☐

If **Yes**, what was the highest level of ARV prescribed? triple or more ☐ 5 dual ☐ 6 mono ☐ 7 N K ☐ 4 month/year started at that level /

Drugs prescribed at that tim

Was this level of ARV treatment maintained to AIDS diagnosis/death without AIDS? Yes ☐ 1 No ☐ 0 N K ☐

If **No** please state reasons

24. Did the patient receive prophylaxis for PCP ***before*** the AIDS diagnosis/death without AIDS? Yes ☐ 1 No ☐ 0 N K ☐

25. Did the patient receive prophylaxis for any other opportunistic infection ***before*** the AIDS diagnosis/death without AIDS? Yes ☐ 1 No ☐ 0 N K ☐

26. Has the patient DIED: Yes ☐ 2 No ☐ 1 If NO, please specify the date of last contact dd/mm/yy / /

If YES, please specify date of death: dd/mm/yy / /

Cause of death (if known):

If patient subsequently dies, please inform CDSC by phone (0181 200 6868 ext 4453) or letter of date and cause of death.

Completed by: ..... **Position:** ..... **Date:** ..... / ..... / .....

Please return the completed form 'in strict medical confidence' to:

Director, CDSC, 61 Colindale Avenue, London NW9 5EQ. Pre-paid address labels supplied.

WHAT TO REPORT ON THIS FORM

1) Any case of AIDS in a patient (alive or dead).

An HIV infected patient is defined as having AIDS if they have one or more of the indicator diseases listed in Q21. The following indicator diseases were added to the list in 1993 and apply only to patients aged 13 or older:

i) Pulmonary tuberculosis
ii) Recurrent pneumonia in a twelve month period
iii) Cervical carcinoma, invasive

The 1993 list of AIDS indicator diseases should be applied to all HIV infected persons alive on or after 1 January 1993,who have not yet been reported as having AIDS.

Most people fulfilling the AIDS case definition will have had a positive HIV antibody test result. In people who have refused an HIV test, definitive methods of diagnosis for AIDS indicator diseases are generally necessary. One of the consultants at the PHLS AIDS and STD Centre at CDSC can be contacted to discuss difficult cases.

2) Any death in an HIV infected person who has died without reaching case definition AIDS.

WHAT TO DO WITH THIS FORM

1). Return the blue copy of the AIDS clinical report form 'in strict medical confidence' to:

The Director, PHLS Communicable Disease Surveillance Centre,
61 Colindale Avenue, London NW9 5EQ. **Use the pre-paid address label supplied.**

2). The pink copy may be sent 'in strict medical confidence' to:

The Physician coordinating AIDS/HIV information in the Health Authority - usually the consultant responsible for communicable disease control, CCDC (see HC[88]64)

3). The white copy may be kept by the reporting clinician as a confidential record.

If not used as suggested above, copies 2 or 3 should be destroyed.

GUIDELINES FOR COMPLETING THE FORM

Q2 - Soundex code of surname
Reporting doctors are encouraged to use the 'soundex code' as an alternative to giving the patient's surname. The soundex code is derived by following a set of eight rules to convert the surname to a first letter followed by three digits. As no surname is unique to a particular code, confidentiality is assured. When the code is used with date of birth and sex, probable duplicate reports can be readily detected and further enquiries made if necessary. The soundex code can be derived by hand, by using programmed calculator or with a short computer program available free of charge from CDSC. For further details of soundex coding contact either CDSC or the physician coordinating AIDS/HIV information in the Health Authority.

Q3 - Dates
For date of birth and date of death please give day/month/year. Elsewhere month/year or approximate date is requested.

Q8 - Place of residence
As boundaries change, postcodes are particularly useful in order to allocate cases to the correct Health Authorities and Local Authorities of residence. In large towns (particularly in London) a full post code is needed because the first part of the post-code may span two or more authorities.

Q12 to 18 - Exposure categories for HIV infection
Reliable information about a patient's risk factors for HIV infection is especially important. Where risk factor information is not recorded in the patient's clinical notes, the AIDS clinical report form should be completed by a physician with personal knowledge of the patient. The patient may need to be re-interviewed.

Q21 - AIDS indicator diseases
Tick all the AIDS indicator diseases (see list) which have been diagnosed. For each disease give month and year when positive result of the definitive diagnostic method became available or when the criteria for a presumptive diagnosis were met.

Unless otherwise stated, ticking a box for an indicator disease listed confirms that the disease was diagnosed by a definitive diagnostic method or that all the appropriate criteria for a presumptive diagnosis were met (see diagnostic criteria in panel opposite).

AIDS INDICATOR DISEASE*	Definitive or Presumptive	Definitive diagnostic method **or** presumptive diagnostic criteria.
Bacterial infections, multiple or recurrent in a child aged less than 13 years	Definitive	culture, antigen detection, CSF microscopy.
Candidiasis, trachea, bronchi or lungs	Definitive	gross inspection at endoscopy/post-mortem or by microscopy (histology or cytology).
Candidiasis of oesophagus	Definitive	gross inspection at endoscopy/post-mortem or by microscopy (histology or cytology).
	Presumptive	recent onset retrosternal pain on swallowing or radiological evidence and confirmed oral or pharyngeal candidiasis.
Cervical carcinoma, invasive	Definitive	histology.
Coccidioidomycosis, disseminated or extrapulmonary	Definitive	microscopy, culture, or antigen detection in affected tissue.
Cryptococcosis, extrapulmonary	Definitive	microscopy, culture, or antigen detection in affected tissue.
Cryptosporidiosis, with diarrhoea for over 1 month	Definitive	stool microscopy.
Cytomegalovirus retinitis	**Presumptive**	loss of vision and characteristic appearance on serial ophthalmoscopy, progressing over several months.
Cytomegalovirus disease (onset after age 1 month) not in liver, spleen or nodes	Definitive	culture lung tissue, microscopy (histology or cytology), antigen or nucleic acid detection.
Encephalopathy (dementia) due to HIV	Definitive	HIV infection and disabling cognitive and/or motor dysfunction, or milestone loss in a child, with no other causes by CSF examination, brain imaging or post-mortem.
Herpes simplex: ulcers for 1 month or bronchitis, pneumonitis, oesophagitis (onset after age 1 month)	Definitive	microscopy, culture, or antigen detection in affected tissue.
Histoplasmosis, disseminated or extrapulmonary	Definitive	microscopy, culture, or antigen detection in affected tissue.
Isosporiasis, with diarrhoea for over 1 month	Definitive	microscopy (histology or cytology).
Kaposi's sarcoma	Definitive	microscopy (histology or cytology).
	Presumptive	characteristic erythematous/violaceous plaque-like lesion on skin or mucous membrane.
Lymphoid interstitial pneumonia and/or pulmonary lymphoid hyperplasia in a child aged less than 13 years	Definitive	microscopy (histology or cytology).
	Presumptive	diffuse bilateral reticulonodular pulmonary interstitial infiltrates for over 2 months and no pathogen identified and no antibiotic response.
Lymphoma: Burkitt's or immunoblastic or primary in brain*	Definitive	microscopy (histology or cytology).
Mycobacteriosis disseminated* (including extrapulmonary TB)	Definitive	culture.
	Presumptive	AFB (species not identified by culture) on microscopy of stool specimen or normally sterile body fluid/tissue, not lungs, skin, cervical or hilar nodes.
Mycobacteriosis pulmonary tuberculosis*	Definitive	culture or other definitive demonstration of *M.tuberculosis* infection.
	Presumptive	clinical diagnosis, with or without AFB on microscopy, resulting in initiation of anti-TB therapy.
Pneumocystis carinii pneumonia	Definitive	microscopy (histology or cytology).
	Presumptive	recent onset dyspnoea on exertion or dry cough, and diffuse bilateral interstitial infiltrates on CXR and pO2 <70mm Hg (9.3kPa) and no evidence of bacterial pneumonia.
Pneumonia recurrent within a 12 mth period	Definitive	two episodes proven microbiologically.
	Presumptive	CXR or clinical diagnosis.
Progressive multifocal leukoencephalopathy	Definitive	electron microscopy, antigen detection in brain or urine, antibody in serum or CSF.
Salmonella (non-typhoid) septicaemia, recurrent	Definitive	culture.
Toxoplasmosis of brain onset after age 1 month	Definitive	microscopy (histology or cytology), mouse inoculation, tissue culture.
	Presumptive	recent onset focal neurological abnormality or reduced level of consciousness, and mass effect lesion on scan, and serological evidence or specific therapy response.
Wasting syndrome due to HIV	Definitive	weight loss (over 10% baseline) with no other cause, and 30 days or more of either diarrhoea or weakness with fever.

*Full case definition and notes on AIDS indicator diseases for neoplasms, mycobacteriosis and indicator diseases in children are available from CDSC.

1993 CLASSIFICATION SYSTEM FOR HIV RELATED DISEASE

Category A: Acute (primary) HIV infection or Asymptomatic HIV infection or Persistent Generalized Lymphadenopathy.

Category B: Symptomatic with conditions other than those included in categories A or C attributed to HIV infection or which are indicative of a defect in cell mediated immunity.

For example:-

Bacillary angiomatosis – Candidiasis, oropharyngeal (thrush) – Candidiasis, vulvovaginal; persistent, frequent, or poorly responsive to therapy – Cervical dysplasia (moderate or severe)/cervical carcinoma in situ – Constitutional symptoms, such as fever (38.5° C) or diarrhoea lasting > 1 month – Hairy leukoplakia, oral – Herpes zoster (shingles), involving at least two distinct episodes or more than one dermatome – Idiopathic thrombocytopenic purpura – Listeriosis – Pelvic inflammatory disease – Peripheral neuropathy.

Category C: Clinical conditions listed in the AIDS surveillance case definition presumptively or definitively diagnosed(see above).

B.2 Clinician HIV Reporting forms

Table B.3. Data collected on CHR forms

	Form number		
Data field	**C0001**	**C2**	**C3**
Confidential report of:			
Newly diagnosed HIV infection	X	X	
and/or AIDS diagnosis	X	X	
and/or death without AIDS in an HIV-infected person	X	X	
First UK diagnosis of HIV infection			X
and/or first UK diagnosis of AIDS			X
and/or death without AIDS in an HIV-infected person			X
To be completed on all reports:			
Reporting consultant	X	X	X
Hospital/centre	X	X	X
Department/ward		X	X
Telephone number	X	X	X
Patient soundex or surname	X	X	X
Initial(s)	X	X	X
Date of birth	X	X	X
Sex	X	X	X
Clinic/Hospital patient identifying number	X	X	X
If in UK temporarily, usual country of residence	X	X	X
Country of birth	X	X	X
If not UK, year of first arrival in UK	X	X	X
Ethnicity	X	X	X
Probable route of HIV infection:			
Sex between men	X	X	X
Year of first sex between men	X	X	X
In last year, estimated no. of male partners	X	X	X
In last year, estimated no. of female partners	X	X	X
Does the patient believe himself to be infected through oral sex		X	X
Injecting drug use	X	X	X
Year first injected	X	X	X
Year last injected	X	X	X
Ever used needle exchange	X		
Sex between men and women (had sex with)	X	X	
Sex between men and women (patient probably infected by)			X
Bisexual male	X	X	X
Injecting drug user	X	X	X
Partner presumed heterosexually infected	X	X	X
If yes, partner's likely country of infection	X	X	X
Other (please specify)	X	X	X
Not known	X	X	X
Comment field for other information	X	X	X
Patient presumed infected in the UK	X	X	X
If no, in which country(ies)	X	X	X
Previously tested		X	X
First ever negative HIV test (date and place)	X	X	
Last negative HIV test (date and place)	X	X	X
First positive test (date and place)	X	X	X
HIV-1	X		
HIV-2	X		

HIV+ (type not known)	X		
Pregnant at diagnosis	X	X	X[1]
If yes, estimated date of delivery	X	X	X[1]
Previous live births since 1980 (year and country of each)	X	X	X
Complete only for HIV diagnoses made after 01/01/2000	X	X	
Patients not previously diagnosed HIV positive in the UK			X
HIV diagnosing laboratory	X	X	X
Date of current diagnosis		X	X
Any evidence of HIV-2		X	X
Date of laboratory report (mm/yy)	X		
Reason for test	X	X	X
Pregnant at diagnosis			X[1]
If yes, estimated date of delivery			X[1]
Ever attended STD clinic before HIV diagnosis	X	X	
If yes, year of first attendance	X	X	
Symptoms (at HIV diagnosis, pre-treatment)	X	X	X
CD4 count (at HIV diagnosis, pre-treatment)	X	X	X
Date of CD4 count (at HIV diagnosis, pre-treatment)	X	X	X
Viral load (at HIV diagnosis, pre-treatment)	X	X	X
Date of viral load (at HIV diagnosis, pre-treatment)	X	X	X
Complete for AIDS diagnosis, or death without AIDS in a HIV+ person:			
AIDS indicator disease	X	X	X
Definitive or presumptive diagnosis	X	X	X
Date of diagnosis (mm/yy)	X	X	X
Any non-HIV related cause of immunodeficiency	X	X	X
Was patient ever on ARV therapy (before AIDS/death without AIDS)	X	X	X
If no, reason		X	X
If yes, approximate duration of pre-AIDS/death without AIDS treatment			X
Date stopped ARV therapy (mm/yy)			X
Reason stopped			X
Highest level of ARV prescribed (before AIDS/death without AIDS)	X	X	
Date started at that level (mm/yy)	X	X	
ARV drugs prescribed at that time	X	X	
Was this level of ARV maintained to AIDS/death without AIDS	X	X	
If yes, factors contributing to compliance	X	X	
If no, reasons	X	X	
Did the patient receive prophylaxis for PCP	X	X	
Did the patient receive prophylaxis for any other opportunistic infection	X	X	
Complete for all reports:			
Has the patient died	X	X	X
If yes, date of death (dd/mm/yy)	X	X	X
Cause of death	X	X	X
If no, date of last contact	X	X	X
Completed by	X	X	X
Position	X	X	X
Date of report	X	X	X
Updated by	X	X	X
Position	X	X	X
Date of report	X	X	X

[1] moved from section for all reports to section for patients not previously diagnosed HIV positive in the UK

Form B.3a. ‘Form C0001’ – ‘HIV, AIDS or HIV Death report’

HIV, AIDS or HIV Death Report

IN STRICT MEDICAL CONFIDENCE

(for those aged 15 years or over at diagnosis)

Clinicians are requested to use these forms to report:-

1) **all first diagnoses of HIV infection made in the UK after the beginning of 2000**
2) **all previously unreported AIDS cases**
3) **all deaths in HIV infected people who have not had an AIDS indicator disease diagnosed**

New diagnoses of HIV infection, or of AIDS, in patients aged less than 15 years are reported through the British Paediatric Surveillance Unit to the Institute of Child Health (London):Telephone 020-7829-8686 for details.

Guidelines for completing the form

If after diagnosis the patient is referred elsewhere for HIV care please forward a copy of this form for completion by the clinician they have been referred to.
Whenever possible please try to report within three months of HIV diagnosis, even if information is incomplete at that time.

Section A: to be completed for all reports
Section B: to be completed for HIV infections diagnosed after 01/01/2000
Section C: to be completed for any HIV infected patient who has developed AIDS or has died without having had an AIDS defining condition
Section D: to be completed for all reports

When an HIV infection diagnosed after 01/01/2000 is reported the second copy of the form may be kept in the patient’s notes. Section C can then be completed and section D updated and returned to CDSC if the patient subsequently develops AIDS or dies.

Dates: please give dates of birth and death in full. For other dates month and year is usually sufficient.

Section A:

Soundex code of surname The ‘soundex’ coding uses a set of eight rules to convert the surname to its first letter followed by three digits. Its use protects confidentiality since no code is unique to a particular surname, but when used with date of birth and sex likely duplicate reports can be readily recognised. The code can be derived by following the rules from a printed sheet, or by using a short computer programme available free from CDSC. For further details or assistance please contact CDSC.

Probable route of HIV infection

Reliable information about the route by which the patient is believed to have been infected, and of the patient’s exposure within that risk category is especially important. Establishing the information may involve re-interviewing the patient.

HIV test History

This seeks where possible to define the period during which infection occurred and to establish whether there were prior opportunities for one-to-one prevention advice.

Pregnancy History

Information about the fertility pattern of HIV positive women is important for interpreting the unlinked anonymous testing programme results.

Section B:

The information asked for in this section will be used to establish the stage of disease progression at which HIV diagnosis has been made.

Section C:

Death without AIDS: This is to be interpreted as death in an HIV infected patient who has never had an AIDS defining condition diagnosed.

AIDS indicator diseases: Tick the AIDS indicator disease(s) diagnosed. For each give as appropriate the month and year of the definitive diagnosis or when the criterion for a presumptive diagnosis was met (see panel).

Treatment Details: Details of treatment **before** AIDS (or death without AIDS). Details of post-AIDS treatment are not sought as the reporting is intended to capture information at the time of AIDS diagnosis. The reporting will be used to monitor the effectiveness of HIV diagnosis and treatment in postponing AIDS, and also whether people on treatment are dying without having had an AIDS defining condition.

NEWLY DIAGNOSED HIV INFECTION (Sections A, B & D) ☐
and/or **AIDS DIAGNOSIS** (Sections A, C & D) ☐
and/or **DEATH WITHOUT AIDS IN AN HIV INFECTED PERSON** (Sections A,C & D) ☐

CONFIDENTIAL REPORT

A: TO BE COMPLETED ON ALL REPORTS

Reporting consultant

Hospital/Centre Tel.

PATIENT DETAILS

Soundex code/surname Initials DOB / / Sex M ☐ F ☐ Clinic/Hosp No

If in UK temporarily, usual country of residence

Country of birth If not UK, year of first arrival in UK

Ethnicity: White ☐ 1 Black Caribbean ☐ 5 Black African ☐ 6 Black Other ☐ 7 Indian/Pakistani/Bangladeshi ☐ 8 Other/mixed ☐ 4 NK ☐ 9

PROBABLE ROUTE OF HIV INFECTION:-PLEASE COMPLETE RELEVANT SECTION(S) AND/OR COMMENT BOX

SEX BETWEEN MEN ☐

Year of first sex between men In last year estimated no. of: Male partners ☐ Female partners ☐

INJECTING DRUG USE ☐

Year first injected Year last injected Ever used needle exchange: Yes ☐ No ☐ NK ☐

SEX BETWEEN MEN AND WOMEN ☐

Has **patient** had sex with: bisexual male Yes ☐ No ☐ NK ☐ Injecting drug user Yes ☐ No ☐ NK ☐

partner presumed heterosexually infected Yes ☐ No ☐ NK ☐

if YES **partner's** likely country(ies) of infection

OTHER ☐ Please specify

NOT KNOWN ☐

PATIENT PRESUMED INFECTED IN UK: Yes ☐ No ☐ NK ☐ If NO: in which country(ies)

HIV TEST HISTORY: (if known)

	Date	Where
first ever **negative** test	/ /	
last **negative** test	/ /	
first **positive** test	/ /	

HIV-1: Pos ☐ Neg ☐ **HIV-2:** Pos ☐ Neg ☐ **HIV+** Type NK ☐

PREGNANCY HISTORY (since 1980):

Pregnant at diagnosis: Yes ☐ No ☐ NK ☐ If YES e.d.d / /

Previous live births: Yes ☐ No ☐ NK ☐ If YES:

year Country

year Country

year Country (cont. in COMMENT)

COMMENT Please give any other relevant details not covered elsewhere, particularly of route of HIV infection e.g. history of transfusion (when, why, where)

B: COMPLETE ONLY FOR HIV DIAGNOSES MADE AFTER 01/01/2000

HIV DIAGNOSING LABORATORY

DATE OF LABORATORY REPORT mm/yy /

Any evidence of date of infection e.g. seroconversion illness, particular incident, laboratory evidence etc

REASON FOR TEST:

Symptoms ☐ 1 Known positive partner ☐ 2 Risky behaviour ☐ 3 Blood donor ☐ 4 Antenatal ☐ 5 Insurance/Visa screen ☐ 6

Confirmation of known positive ☐ 7 Other (specify) NK ☐ 9

EVER ATTENDED ANY STD CLINIC BEFORE HIV DIAGNOSIS: Yes ☐ No ☐ NK ☐ If YES, year of first attendance

Section B continued.....

BASELINE AT HIV DIAGNOSIS (PRE TREATMENT):

SYMPTOMS: AIDS defining ☐ | HIV related (not AIDS defining) ☐ | None/not HIV related ☐

CD4 COUNT ____ (cells/µl) mm/yy / Not yet done ☐

VIRAL LOAD ____ (copies/ml) mm/yy / Not yet done ☐

CDSC use only C0001

RepDate

CHIVRep C

AIDS/HIV death

C: COMPLETE FOR AIDS DIAGNOSIS, OR DEATH WITHOUT AIDS IN AN HIV INFECTED PERSON

Please tick the AIDS indicator disease(s) diagnosed (unless HIV death report) and give date of diagnosis

AIDS INDICATOR DISEASE	DIAGNOSIS definitive		DATE mm/yy
Candidiasis: trachea, bronchi or lungs	☐ 2	NA	/
Candidiasis: oesophageal	☐ 3	☐	/
Cervical carcinoma, invasive	☐ 4	NA	/
Coccidioidomycosis: extrapulmonary	☐ 5	NA	/
Cryptococcosis: extrapulmonary	☐ 6	NA	/
Cryptosporidiosis with diarrhoea for > 1 month	☐ 7	NA	/
Cytomegalovirus retinitis	☐ 8	☐	/
CMV disease not in liver, spleen or nodes	☐ 9	NA	/
Encephalopathy (dementia) due to HIV	☐ 10	NA	/
Herpes simplex: ulcers > 1 month, or bronchitis, pneumonitis, oesophagitis	☐ 11	NA	/
Histoplasmosis: disseminated/extrapulmonary	☐ 12	NA	/
Isosporiasis: with diarrhoea for > 1 month	☐ 13	NA	/
Kaposi's sarcoma	☐ 14	☐	/

AIDS INDICATOR DISEASE	DIAGNOSIS definitive		DATE mm/yy
Lymphoma, Burkitt's or equivalent term	☐ 16	NA	/
Lymphoma, immunoblastic or equivalent term	☐ 17	NA	/
Lymphoma, primary in brain	☐ 18	NA	/
Mycobacterium avium: extrapulmonary	☐ 19	☐	/
M. tuberculosis: pulmonary	☐ 20	☐	/
M. tuberculosis: extrapulmonary	☐ 21	☐	/
Mycobacterium of other or unidentified species, disseminated	☐ 22	☐	/
Pneumocystis carinii pneumonia	☐ 23	☐	/
Pneumonia: recurrent within 12 months	☐ 24	☐	/
Progressive multifocal leukoencephalopathy	☐ 25	NA	/
Salmonella septicaemia, recurrent	☐ 26	NA	/
Toxoplasmosis of the brain	☐ 27	☐	/
Wasting syndrome due to HIV	☐ 28	NA	/

Please state any non-HIV related cause of immunodeficiency

TREATMENT: BEFORE Diagnosis of AIDS, or Death without AIDS:

Was the patient ever on anti retroviral therapy? Yes ☐ No ☐ NK ☐ NA ☐

If YES: what was the highest level of ARV prescribed? dual ☐ triple ☐ other ☐ NK ☐

month/year started at that level /

ARV drugs prescribed at that time

Was this level of ARV treatment maintained to AIDS diagnosis/death without AIDS? Yes ☐ No ☐ NK ☐

If YES factors contributing to progression: poor compliance ☐ viral resistance ☐ other ____

If NO please state reasons: treatment not tolerated ☐ patient choice ☐ other ____

Did the patient receive prophylaxis for PCP? Yes ☐ No ☐ NK ☐ NA ☐

Did the patient receive prophylaxis for any other opportunistic infection? Yes ☐ No ☐ NK ☐ NA ☐

D: COMPLETE FOR ALL REPORTS

Has the patient DIED? Yes ☐ No ☐ NK ☐ If YES, please give date of death: dd/mm/yy / /

Cause of death(if known) ____

If patient is not dead please give date of last contact mm/yy /

If patient subsequently dies, please inform CDSC by phone (020-8200 6868 ext. 4453) or letter stating date and cause of death

Completed by: ____ **Position:** ____ **Date:** / /

Updated by: ____ Position: ____ Date: / /

Please return completed form, "in strict medical confidence", to:

Director, CDSC, 61 Colindale Avenue, London NW9 5EQ

Pre-paid labels supplied on request

AIDS INDICATOR DISEASE	Definitive or Presumptive	Definitive diagnostic method or presumptive diagnostic criteria
*Bacterial infections, multiple or recurrent in a child aged less than 13 years	Definitive	Culture, antigen detection, CSF microscopy
Candidiasis, trachea, bronchi or lungs	Definitive	Gross inspection at endoscopy/post-mortem or by microscopy (histology or cytology).
Candidiasis of oesophagus	Definitive **Presumptive**	Gross inspection at endoscopy/post-mortem or by microscopy (histology or cytology). Recent onset retrosternal pain on swallowing or radiological evidence and confirmed oral or pharyngeal candidiasis.
**Cervical carcinoma, invasive	Definitive	Histology
Coccidioidomycosis, disseminated or extrapulmonary	Definitive	Microscopy, culture, or antigen detection in affected tissue.
Cryptococcosis, extrapulmonary	Definitive	Microscopy, culture, or antigen detection in affected tissue.
Cryptosporidiosis, with diarrhoea for over 1 month	Definitive	Stool microscopy
Cytomegalovirus retinitis	**Presumptive**	Loss of vision and characteristic appearance on serial ophthalmoscopy, progressing over several months.
Cytomegalovirus disease (onset after age 1 month)not in liver, spleen or nodes	Definitive	Culture lung tissue, microscopy (histology or cytology), antigen or nucleic acid detection.
Encephalopathy (dementia) due to HIV	Definitive	HIV infection and disabling cognitive and/or motor dysfunction,**or milestone loss in a child*, with no other causes by CSF examination, brain imaging or post-mortem.
Herpes simplex: ulcers for 1 month or bronchitis, pneumonitis, oesophagitis (onset after age 1 month)	Definitive	Microscopy, culture, or antigen detection in affected tissue.
Histoplasmosis, disseminated or extrapulmonary	Definitive	Microscopy, culture, or antigen detection in affected tissue.
Isoporiasis, with diarrhoea for over 1 month	Definitive	Microscopy (histology or cytology).
Kaposi's sarcoma	Definitive **Presumptive**	Microscopy (histology or cytology). Characteristic erythematous/violaceous plaque-like lesion on skin or mucous membrane.
*Lymphoid interstitial pneumonia and/or pulmonary lymphoid hyperplasia in a child aged less than 13 years	Definitive **Presumptive**	Microscopy (histology or cytology) Diffuse bilateral reticulonodular pulmonary interstitial infiltrates for over 2 months and no pathogen identified and no antibiotic response
Lymphoma: Burkitt's or immunoblastic or primary in brain	Definitive	Microscopy (histology or cytology).
Mycobacteriosis disseminated (including extrapulmonary TB)	Definitive **Presumptive**	Culture. AFB (species not identified by culture) on microscopy of stool specimen or normally sterile body fluid/tissue, not lungs, skin, cervical or hilar nodes.
**Mycobacteriosis pulmonary tuberculosis	Definitive **Presumptive**	Culture or other definitive demonstration of *M.tuberculosis* infection. Clinical diagnosis, with or without AFB on microscopy, resulting in initiation of anti-TB therapy.
Pneumocystis carinii pneumonia	Definitive **Presumptive**	Microscopy (histology or cytology). Recent onset dyspnoea on exertion or dry cough, and diffuse bilateral interstitial infiltrates on CXR and p02<70mm Hg (9.3kPa) and no evidence of bacterial pneumonia.
**Pneumonia recurrent within a 12 month period	Definitive **Presumptive**	Two episodes proven microbiologically. CXR or clinical diagnosis.
Progressive multifocal leukoencephalopathy	Definitive	Electron microscopy, antigen detection in brain and urine, antibody in serum or CSF
Salmonella (non-typhoid) septicaemia, recurrent	Definitive	Culture
Toxoplasmosis of brain onset after age 1 month	Definitive **Presumptive**	Microscopy (histology or cytology), mouse inoculation, tissue culture. Recent onset focal neurological abnormality or reduced level of consciousness, and mass effect lesion on scan, and serological evidence or specific therapy response.
Wasting syndrome due to HIV	Definitive	Weight loss (over 10% baseline) with no other cause, and 30 days or more of either diarrhoea or weakness with fever.

*Applicable only to patients aged < 13 years when diagnosed.
** Applicable only to patients aged >= 13 years diagnosed with these conditions on or after 1st January 1993

What to do with the completed form

1) Return the buff top sheet of the report form, "in strict medical confidence", to: ***The Director, PHLS Communicable Disease Surveillance Centre, 61 Colindale Avenue, London NW9 5EQ*** using the pre-paid labels supplied.
2) The second sheet of the completed form may be kept by the reporting clinician in the patient's notes or elsewhere, and may be copied and used to report AIDS diagnosis, or death without AIDS, to CDSC should either event occur.
3) The third copy may be sent, "in strict medical confidence", to: The physician coordinating HIV/AIDS information in the Health Authority, usually the CCDC

Please ring the PHLS AIDS and STD Centre on 020-8200-6868, ext 4406 for further copies of this form.

Form B.3b. 'Form C2' – 'HIV, AIDS or HIV Death report'

IN STRICT MEDICAL CONFIDENCE

HIV, AIDS or HIV Death report

(for those aged 15 years or over at diagnosis)

Clinicians are requested to use these forms to report:-

1) **all first diagnoses of HIV infection made in the UK after the beginning of 2000**
2) **all previously unreported first UK diagnoses of AIDS**
3) **all deaths in HIV infected people who have not had an AIDS indicator disease diagnosed**

New diagnoses of HIV infection, or of AIDS, in patients aged less than 15 years are reported through the British Paediatric Surveillance Unit to the Institute of Child Health (London):Telephone 020-7829-8686 for details.

Guidelines for completing the form

If after diagnosis the patient is referred elsewhere for HIV care please forward a copy of this form for completion by the clinician they have been referred to.
Whenever possible please try to report within three months of HIV diagnosis, even if information is incomplete at that time.

Section A: to be completed for all reports
Section B: to be completed for HIV infections diagnosed after 01/01/2000
Section C: to be completed for any HIV infected patient who has an AIDS indicator disease diagnosed in the UK or has died without having had an AIDS defining condition
Section D: to be completed for all reports

When an HIV infection diagnosed after 01/01/2000 is reported the second copy of the form may be kept in the patient's notes. Section C can then be completed and section D updated and returned to CDSC if the patient subsequently develops AIDS or dies without having had an AIDS defining condition.

Dates: please give dates of birth and death in full. For other dates month and year are usually sufficient.

Section A:

Soundex code of surname The 'soundex' coding uses a set of eight rules to convert the surname to its first letter followed by three digits. Its use protects confidentiality since no code is unique to a particular surname, but when used with date of birth and sex likely duplicate reports can be readily recognised. The code can be derived by following the rules from a printed sheet, or by using a short computer programme available free from CDSC. For further details or assistance please contact CDSC.

Probable route of HIV infection

Reliable information about the route by which the patient is believed to have been infected, and of the patient's exposure within that risk category is especially important. Establishing the information may involve re-interviewing the patient.

HIV test History

This seeks where possible to define the period during which infection occurred and to establish whether there were prior opportunities for one-to-one prevention advice.

Pregnancy History

Information about the fertility pattern of HIV positive women is important for interpreting the unlinked anonymous testing programme results.

Section B:

The information asked for in this section will be used to establish the stage of disease progression at which HIV diagnosis has been made.

Section C:

Death without AIDS: This is to be interpreted as death in an HIV infected patient who has never had an AIDS defining condition diagnosed.

AIDS indicator diseases: Tick the AIDS indicator disease(s) diagnosed. For each give as appropriate the month and year of the definitive diagnosis or when the criterion for a presumptive diagnosis was met (see panel).

Treatment Details: Details of treatment **before** AIDS (or death without AIDS). Details of post-AIDS treatment are not sought as the reporting is intended to capture information at the time of AIDS diagnosis. The reporting will be used to monitor the effectiveness of HIV diagnosis and treatment in postponing AIDS, and also whether people on treatment are dying without having had an AIDS defining condition.

CONFIDENTIAL REPORT OF:

☐ **NEWLY DIAGNOSED HIV INFECTION** (Sections A, B & D)
and/or ☐ **AIDS DIAGNOSIS** (Sections A, C & D)
and/or ☐ **DEATH WITHOUT AIDS IN AN HIV INFECTED PERSON** (Sections A, C & D)

CDSC use only C2
RepDate
CHIVRep C
AIDS/HIV death

A: TO BE COMPLETED ON ALL REPORTS

Hospital/Centre | Dept./Ward
Reporting Consultant | Tel.

PATIENT DETAILS

Soundex code/surname | Initials | DOB / / | Sex M F | Clinic/Hosp No

If in UK temporarily, usual country of residence

Country of birth | If not UK, year of first arrival in UK

Ethnicity: White 1 | Black Caribbean 5 | Black African 6 | Black Other 7 | Indian/Pakistani/Bangladeshi 8 | Other/mixed 4 | NK 9

PROBABLE ROUTE OF HIV INFECTION: PLEASE COMPLETE RELEVANT SECTION(S) AND/OR COMMENT BOX

PATIENT PRESUMED INFECTED IN UK: Yes | No | NK | If NO: in which country(ies)

☐ **SEX BETWEEN MEN**

Year of first sex between men | In last year estimated no. of: Male partners | Female partners
Does the patient believe himself to have been infected through oral sex? Yes | No | NK

☐ **INJECTING DRUG USE**

Year first injected | Year last injected

☐ **SEX BETWEEN MEN AND WOMEN**

Has **patient** had sex with: bisexual male Yes | No | NK | Injecting drug user Yes | No | NK
partner presumed heterosexually infected Yes | No | NK
if YES **partner's** likely country(ies) of infection

☐ **OTHER** Please specify
☐ **NOT KNOWN**

HIV TEST HISTORY PRIOR TO THIS DIAGNOSIS

previously tested ? Yes | No | NK

first ever **negative** test / /
Where?

last **negative** test / /
Where?

earliest **positive** test / /
Where?

COMMENT Please give any relevant details not covered elsewhere, particularly of route of HIV infection e.g. history of transfusion (when, why, where) and of evidence of recent seroconvertion

B: COMPLETE ONLY FOR HIV DIAGNOSES MADE AFTER 01/01/2000

HIV DIAGNOSING LABORATORY
DATE OF CURRENT DIAGNOSIS / /
Any evidence of HIV-2? Yes | No | NK | If yes, give details in **COMMENT** above

REASON FOR TEST:

Symptoms 1 | Known positive partner 2 | Risky behaviour 3 | Blood donor 4 | Antenatal 5 | Insurance/Visa screen 6
Confirmation of known positive 7 | Other (specify) | NK 9

EVER ATTENDED ANY STD CLINIC BEFORE HIV DIAGNOSIS: Yes | No | NK | If YES, year of first attendance

Section B continued.....

PREGNANCY HISTORY (since 1980):

Pregnant at diagnosis: Yes ☐ No ☐ NK ☐ e.d.d / /

Previous live births: Yes ☐ No ☐ NK ☐ If YES:

year ☐ Country ☐

year ☐ Country ☐

year ☐ Country ☐ (cont. in **COMMENT**)

BASELINE AT HIV DIAGNOSIS (PRE TREATMENT):

SYMPTOMS : Seroconversion ☐ 1 None/not HIV related ☐ 2

HIV related (not AIDS defining) ☐ 3 AIDS defining ☐ 4

CD4 COUNT ☐ (cells/µl) / / Not yet done ☐

VIRAL LOAD ☐ (copies/ml) / / Not yet done ☐

C: COMPLETE FOR AIDS DIAGNOSIS, OR DEATH WITHOUT AIDS IN AN HIV INFECTED PERSON

Please tick the AIDS indicator disease(s) diagnosed (unless HIV death report) and give date of diagnosis

AIDS INDICATOR DISEASE	DIAGNOSIS definitive	DIAGNOSIS presumptive	DATE mm/yy
Candidiasis: trachea, bronchi or lungs	☐ 2	NA	/
Candidiasis: oesophageal	☐ 3	☐	/
Cervical carcinoma, invasive	☐ 4	NA	/
Coccidioidomycosis: extrapulmonary	☐ 5	NA	/
Cryptococcosis: extrapulmonary	☐ 6	NA	/
Cryptosporidiosis with diarrhoea for > 1 month	☐ 7	NA	/
Cytomegalovirus retinitis	☐ 8	☐	/
CMV disease not in liver, spleen or nodes	☐ 9	NA	/
Encephalopathy (dementia) due to HIV	☐ 10	NA	/
Herpes simplex: ulcers > 1 month, or bronchitis, pneumonitis, oesophagitis	☐ 11	NA	/
Histoplasmosis: disseminated/extrapulmonary	☐ 12	NA	/
Isosporiasis: with diarrhoea for > 1 month	☐ 13	NA	/
Kaposi's sarcoma	☐ 14	☐	/

AIDS INDICATOR DISEASE	DIAGNOSIS definitive	DIAGNOSIS presumptive	DATE mm/yy
Lymphoma, Burkitt's or equivalent term	☐ 16	NA	/
Lymphoma, immunoblastic or equivalent term	☐ 17	NA	/
Lymphoma, primary in brain	☐ 18	NA	/
Mycobacterium avium: extrapulmonary	☐ 19	☐	/
M. tuberculosis: pulmonary	☐ 20	☐	/
M. tuberculosis: extrapulmonary	☐ 21	☐	/
Mycobacterium of other or unidentified species, disseminated	☐ 22	☐	/
Pneumocystis carinii pneumonia	☐ 23	☐	/
Pneumonia: recurrent within 12 months	☐ 24	☐	/
Progressive multifocal leukoencephalopathy	☐ 25	NA	/
Salmonella septicaemia, recurrent	☐ 26	NA	/
Toxoplasmosis of the brain	☐ 27	☐	/
Wasting syndrome due to HIV	☐ 28	NA	/

Please state any non-HIV related cause of immunodeficiency ☐

TREATMENT: BEFORE Diagnosis of AIDS, or Death without AIDS:

Was the patient ever on anti retroviral therapy? Yes ☐ No ☐ NK ☐ NA ☐

If no: Reason ☐

If YES: what was the highest level of ARV prescribed? dual ☐ triple ☐ other ☐ NK ☐ date started at that level /

ARV drugs prescribed at that time ☐

Was this level of ARV treatment maintained to AIDS diagnosis/death without AIDS? Yes ☐ No ☐ NK ☐

If YES factors contributing to progression: poor compliance ☐ viral resistance ☐ other ☐

If NO please state reasons: treatment not tolerated ☐ patient choice ☐ other ☐

and date stopped /

Did the patient receive prophylaxis for PCP? Yes ☐ No ☐ NK ☐ NA ☐

Did the patient receive prophylaxis for any other opportunistic infection? Yes ☐ No ☐ NK ☐ NA ☐

D: COMPLETE FOR ALL REPORTS

Has the patient DIED? Yes ☐ No ☐ NK ☐ If YES, please give date of death: dd/mm/yy / /

Cause of death(if known) ☐

If patient is not dead please give date of last contact mm/yy /

If patient subsequently dies, please inform CDSC by phone (020-8200 6868 ext. 4453) or letter stating date, place and cause of death

Completed by: ☐ **Position:** ☐ **Date:** / /

Updated by: ☐ Position: ☐ Date: / /

Please return completed form, 'in strict medical confidence', to:

Director, CDSC, 61 Colindale Avenue, London NW9 5EQ marked 'Attention HARS' ☐ **Pre-paid labels supplied on request**

AIDS INDICATOR DISEASE	Definitive or Presumptive	Definitive diagnostic method or presumptive diagnostic criteria
*Bacterial infections, multiple or recurrent in a child aged less than 13 years	Definitive	Culture, antigen detection, CSF microscopy
Candidiasis, trachea, bronchi or lungs	Definitive	Gross inspection at endoscopy/post-mortem or by microscopy (histology or cytology).
Candidiasis of oesophagus	Definitive **Presumptive**	Gross inspection at endoscopy/post-mortem or by microscopy (histology or cytology). Recent onset retrosternal pain on swallowing or radiological evidence and confirmed oral or pharyngeal candidiasis.
**Cervical carcinoma, invasive	Definitive	Histology
Coccidioidomycosis, disseminated or extrapulmonary	Definitive	Microscopy, culture, or antigen detection in affected tissue.
Cryptococcosis, extrapulmonary	Definitive	Microscopy, culture, or antigen detection in affected tissue.
Cryptosporidiosis, with diarrhoea for over 1 month	Definitive	Stool microscopy
Cytomegalovirus retinitis	**Presumptive**	Loss of vision and characteristic appearance on serial ophthalmoscopy, progressing over several months.
Cytomegalovirus disease (onset after age 1 month) not in liver, spleen or nodes	Definitive	Culture lung tissue, microscopy (histology or cytology), antigen or nucleic acid detection.
Encephalopathy (dementia) due to HIV	Definitive	HIV infection and disabling cognitive and/or motor dysfunction,*or *milestone loss in a child*, with no other causes by CSF examination, brain imaging or post-mortem.
Herpes simplex: ulcers for 1 month or bronchitis, pneumonitis, or oesophagitis (onset after age 1 month)	Definitive	Microscopy, culture, or antigen detection in affected tissue.
Histoplasmosis, disseminated or extrapulmonary	Definitive	Microscopy, culture, or antigen detection in affected tissue.
Isoporiasis, with diarrhoea for over 1 month	Definitive	Microscopy (histology or cytology).
Kaposi's sarcoma	Definitive **Presumptive**	Microscopy (histology or cytology). Characteristic erythematous/violaceous plaque-like lesion on skin or mucous membrane.
*Lymphoid interstitial pneumonia and/or pulmonary lymphoid hyperplasia in a child aged less than 13 years	Definitive **Presumptive**	Microscopy (histology or cytology) Diffuse bilateral reticulonodular pulmonary interstitial infiltrates for over 2 months and no pathogen identified and no antibiotic response
Lymphoma: Burkitt's or immunoblastic or primary in brain	Definitive	Microscopy (histology or cytology).
Mycobacteriosis: disseminated (including extrapulmonary TB)	Definitive **Presumptive**	Culture. AFB (species not identified by culture) on microscopy of stool specimen or normally sterile body fluid/tissue, not lungs, skin, cervical or hilar nodes.
**Mycobacteriosis: pulmonary tuberculosis	Definitive **Presumptive**	Culture or other definitive demonstration of *M.tuberculosis* infection. Clinical diagnosis, with or without AFB on microscopy, resulting in initiation of anti-TB therapy.
Pneumocystis carinii pneumonia	Definitive **Presumptive**	Microscopy (histology or cytology). Recent onset dyspnoea on exertion or dry cough, and diffuse bilateral interstitial infiltrates on CXR and p02<70mm Hg (9.3kPa) and no evidence of bacterial pneumonia.
**Pneumonia recurrent within a 12 month period	Definitive **Presumptive**	Two episodes proven microbiologically. CXR or clinical diagnosis.
Progressive multifocal leukoencephalopathy	Definitive	Electron microscopy, antigen detection in brain and urine, antibody in serum or CSF
Salmonella (non-typhoid) septicaemia, recurrent	Definitive	Culture
Toxoplasmosis of brain onset after age 1 month	Definitive **Presumptive**	Microscopy (histology or cytology), mouse inoculation, tissue culture. Recent onset focal neurological abnormality or reduced level of consciousness, and mass effect lesion on scan, and serological evidence or specific therapy response.
Wasting syndrome due to HIV	Definitive	Weight loss (over 10% baseline) with no other cause, and 30 days or more of either diarrhoea or weakness with fever.

*Applicable only to patients aged < 13 years when diagnosed.
** Applicable only to patients aged ≥ 13 years diagnosed with these conditions on or after 1st January 1993

What to do with the completed form

1) Return the buff top sheet of the report form, 'in strict medical confidence', to: ***The Director, PHLS Communicable Disease Surveillance Centre, 61 Colindale Avenue, London NW9 5EQ*** marking envelope 'Attention HARS' if not using the pre-paid labels supplied.
2) The second sheet of the completed form may be kept by the reporting clinician in the patient's notes or elsewhere, and may be copied and used to report AIDS diagnosis, or death without AIDS, to CDSC should either event occur.
3) The third copy may be kept with the patient's notes as a record of the reporting to CDSC.

Please ring the PHLS AIDS and STD Centre on 020-8200-6868, ext 4406 for further copies of this form.

Form B.3c. ‘Form C3’ – ‘HIV, AIDS or HIV Death report’

IN STRICT MEDICAL CONFIDENCE

HIV, AIDS or HIV Death report

(for those aged 15 years or over at diagnosis)

Clinicians are requested to use these forms to report:-

1) all first diagnoses of HIV infection made in the UK after the beginning of 2000
2) all previously unreported first UK diagnoses of AIDS
3) all deaths in HIV infected people who have not had an AIDS indicator disease diagnosed

*New diagnoses of HIV infection, or of AIDS, in patients **aged less than 15 years** are reported through the British Paediatric Surveillance Unit to the Institute of Child Health (London): Telephone 020-7829-8686 for details.*

GUIDELINES FOR COMPLETING THE FORM

If after diagnosis the patient is referred elsewhere for HIV care please forward a copy of this form for completion by the clinician they have been referred to.

Whenever possible please report within three months of HIV diagnosis, even if information is incomplete at that time.

Section A: to be completed for all reports

Section B: to be completed for **first UK diagnosis of HIV infection***

Section C: to be completed for **any** HIV infected patient who has an AIDS indicator disease diagnosed in the UK or has died without having had an AIDS defining condition

Section D: to be completed for all reports.

** Excluding patients diagnosed before 01/01/2000, and patients who have transferred from elsewhere in the UK.*

When an HIV diagnosis is reported the second copy of the form may be kept in the patients' notes. If the patient subsequently develops AIDS or dies without having had an AIDS defining condition, section C can be completed and section D updated before returning the copy to CDSC.

Section A

Soundex code of surname The 'soundex' coding uses a set of eight rules to convert the surname to its first letter followed by three digits. Its use protects confidentiality since no code is unique to a particular surname, but when used with date of birth and sex likely duplicate reports can be readily recognised. The code can be derived by following the rules from a printed sheet, or by using a short computer programme available free from CDSC. For further details or assistance please contact CDSC.

Probable route of HIV infection
Reliable information about the route by which the patient is believed to have been infected, and of the patient's exposure within that risk category is especially important. Establishing the information may involve re-interviewing the patient.

HIV test history
This seeks where possible to define the period during which infection occurred.

Section B

The information asked for this section will be used to establish the stage of disease progression at which HIV diagnosis has been made.

Section C

Death without AIDS: This is to be interpreted as death in an HIV infected patient who has never had an AIDS defining condition diagnosed.

Aids indicator diseases: Tick the AIDS indicator disease(s) diagnosed. For each give as appropriate the month and year of the definitive diagnosis or when the criterion for a presumptive diagnosis was met (see panel).

Treatment pre AIDS or death without AIDS: Outline information is sought on the use of ART in those progressing to AIDS or pre-AIDS death despite a previous HIV diagnosis.

What to do with the completed form

1) Return the buff top sheet of the report form, 'in strict medical confidence', to: ***The Director, PHLS Communicable Disease Surveillance Centre, 61 Colindale Avenue, London NW9 5EQ*** marking envelope 'Attention HARS' if not using the pre-paid labels supplied.
2) The second sheet of the completed form may be kept by the reporting clinician in the patient's notes or elsewhere, and may be copied and used to report AIDS diagnosis, or death without AIDS, to CDSC should either event occur.
3) The third copy may be kept with the patient's notes as a record of the reporting to CDSC.

CONFIDENTIAL REPORT OF:

- ☐ **FIRST UK DIAGNOSIS OF HIV INFECTION** (Sections A, B & D)
- and/or ☐ **FIRST UK DIAGNOSIS OF AIDS** (Sections A, C & D)
- and/or ☐ **DEATH WITHOUT AIDS IN AN HIV INFECTED PERSON** (Sections A, C & D)

CDSC use only **C3**

Rep Date

CHR No C

AIDS/ DwoA No

A: ALL REPORTS

Hospital/Centre | Dept./Ward

Reporting Consultant | Telephone

PATIENT DETAILS

Soundex code/surname | Initials | DOB / / | Sex M ☐ F ☐ | Clinic/Hosp No

If in UK temporarily, usual country of residence

Country of birth | In not UK, date of first arrival (mm/yy) /

Ethnicity: White ☐ 1 | Black Caribbean ☐ 5 | Black African ☐ 6 | Black Other ☐ 7 | Indian sub-continent ☐ 8 | Other/mixed ☐ 4 | NK ☐ 9

PROBABLE ROUTE OF HIV INFECTION: PLEASE COMPLETE RELEVANT SECTION(S) AND/OR COMMENT BOX

PATIENT PRESUMED INFECTED IN UK: Yes ☐ No ☐ NK ☐ If NO: in which country(ies)

☐ **SEX BETWEEN MEN**

Year of first sex between men | In **last year** estimated no. of: Male partners | Female partners

Does the patient believe himself to have been infected through oral sex? Yes ☐ No ☐ NK ☐

☐ **SEX BETWEEN MEN & WOMEN**

Was **patient** probably infected by: Bisexual male Yes ☐ No ☐ NK ☐ Injecting drug user Yes ☐ No ☐ NK ☐

Heterosexually infected partner Yes ☐ No ☐ NK ☐

If YES **partner's** likely country(ies) of infection

☐ **INJECTING DRUG USE**

Year first injected | Year last injected

☐ **OTHER** Please specify

☐ **NOT KNOWN**

HIV TEST HISTORY PRIOR TO THIS DIAGNOSIS

previously tested ? Yes ☐ No ☐ NK ☐

last **negative** test / /

Where?

earliest **positive** test / /

Where?

COMMENT Relevant details not covered elsewhere, particularly of route of HIV infection e.g. history of transfusion (when, why, where) and of evidence of recent seroconversion

CDSC use only

Key words

B: PATIENTS NOT PREVIOUSLY DIAGNOSED HIV POSITIVE IN THE UK

DATE OF CURRENT DIAGNOSIS / / DIAGNOSING LABORATORY

Laboratory evidence of recent infection? Not investigated ☐ None ☐ Yes: specify

Any evidence of HIV-2? Yes ☐ No ☐ NK ☐ If yes, give details in **COMMENT**

REASON FOR TEST:

Symptoms ☐ 1 Known positive partner ☐ 2 Risky behaviour ☐ 3 Blood donor ☐ 4 Antenatal ☐ 5 Insurance/Visa screen ☐ 6

Confirmation of known positive ☐ 7 Routine GUM screen ☐ 18 Other (specify) NK ☐ 99

Pregnant at diagnosis: Yes ☐ No ☐ NK ☐ e.d.d / /

BASELINE AT HIV DIAGNOSIS (PRE UK TREATMENT):

SYMPTOMS: Seroconversion ☐ 1 None/not HIV related ☐ 2

HIV related (not AIDS defining) ☐ 3 AIDS defining ☐ 4

CD4 COUNT (cells/µl) / / Not yet done ☐

VIRAL LOAD (copies/ml) / / Not yet done ☐

C: AIDS DIAGNOSIS, OR DEATH WITHOUT AIDS IN AN HIV INFECTED PERSON

Tick the AIDS indicator disease(s) diagnosed and give date of diagnosis

AIDS INDICATOR DISEASE	DIAGNOSIS definitive	DIAGNOSIS presumptive	DATE mm/yy
Candidiasis: trachea, bronchi or lungs	☐ 2	NA	/
Candidiasis:oesophageal	☐ 3	☐	/
Cervical carcinoma, invasive	☐ 4	NA	/
Coccidioidomycosis: extrapulmonary	☐ 5	NA	/
Cryptococcosis: extrapulmonary	☐ 6	NA	/
Crytosporidiosis with diarrhoea for > 1 month	☐ 7	NA	/
Cytomegalovirus retinitis	☐ 8	☐	/
CMV disease not in liver, spleen or nodes	☐ 9	NA	/
Encephalopathy (dementia) due to HIV	☐ 10	NA	/
Herpes simplex: ulcers > 1 month, or bronchitis, pneumonitis, oesophagitis	☐ 11	NA	/
Histoplasmosis: disseminated/extrapulmonary	☐ 12	NA	/
Isosporiasis: with diarrhoea for > 1 month	☐ 13	NA	/
Kaposi's sarcoma	☐ 14	☐	/

AIDS INDICATOR DISEASE	DIAGNOSIS definitive	DIAGNOSIS presumptive	DATE mm/yy
Lymphoma, Burkitt's or equivalent term	☐ 16	NA	/
Lymphoma, immunoblastic or equivalent term	☐ 17	NA	/
Lymphoma, primary in brain	☐ 18	NA	/
Mycobacterium avium: extrapulmonary	☐ 19	☐	/
M. tuberculosis: pulmonary	☐ 20	☐	/
M. tuberculosis: extrapulmonary	☐ 21	☐	/
Mycobacterium of other or unidentified species, disseminated	☐ 22	☐	/
Pneumocystis carinii pneumonia	☐ 23	☐	/
Pneumonia: recurrent within 12 months	☐ 24	☐	/
Progressive multifocal leukoencephalopathy	☐ 25	NA	/
Salmonella septicaemia, recurrent	☐ 26	NA	/
Toxoplasmosis of the brain	☐ 27	☐	/
Wasting syndrome due to HIV	☐ 28	NA	/

Please state any non-HIV related cause of immunodeficiency

Did patient have any ART before AIDS/Death without AIDS Yes ☐ No ☐ NK ☐

If **no**, why not? Simultaneous HIV/AIDS diagnosis ☐ Patient choice ☐ Clinical decision ☐ Ineligible for funding ☐

Other

If **yes**, approximate duration of pre-AIDS/Death without AIDS treatment? < 1 month ☐ 2-3 months ☐ > 3 months ☐ NK ☐

Date stopped (if applicable) mm/yy /

Reason stopped (tick any that apply) Drug toxicity ☐ Drug failure ☐ Patient choice ☐ Other

D: ALL REPORTS

Has the patient **DIED**? Yes ☐ No ☐ NK ☐ If YES, **date of death** / /

Cause of death (if known)

If patient is not dead please give date of last contact / /

If patient subsequently dies, please inform CDSC by phone (0208 200 6868 ext. 4453) or letter stating date, place and cause of death

Completed by: **Position:** **Date:** / /

Updated by: Position: Date: / /

Please return completed form, 'in strict medical confidence', to:

Director, CDSC, 61 Colindale Avenue, London NW9 5EQ marked 'Attention HARS'

Pre-paid labels supplied on request

AIDS INDICATOR DISEASE	Definitive or Presumptive	Definite diagnostic method or presumptive diagnostic criteria
Candidiasis of trachea, bronchi or lungs	Definitive	Gross inspection at endoscopy/post-mortem or by microscopy (histology or cytology).
Candidiasis of oesophagus	Definitive *Presumptive*	Gross inspection at endoscopy/post mortem or by microscopy (histology or cytology). *Recent onset restrosternal pain on swallowing or radiological evidence and confirmed oral or pharyngeal candidiasis*
*Cervical carcinoma, invasive	Definitive	Histology
Coccidioidomycosis, disseminated or extrapulmonary	Definitive	Microscopy, culture, or antigen detection in affected tissue.
Cryptococcosis, extrapulmonary	Definitive	Microscopy, culture, or antigen detection in affected tissue.
Cryptosporidiosis, with diarrhoea for over 1 month	Definitive	Stool microscopy
Cytomegalovirus retinitis	*Presumptive*	*Loss of vision and characteristic appearance on serial ophthalmoscopy, progressing over several months.*
Cytomegalovirus disease (onset after age 1 month) not in liver, spleen or nodes	Definitive	Culture lung tissue, microscopy (histology or cytology), antigen or nucleic acid detection.
Encephalopathy (dementia) due to HIV	Definitive	HIV infection and disabling cognitive and/or motor dysfunction,**or milestone loss in a child, with no other causes by CSF examination, brain imaging or post-mortem.
Herpes simplex: ulcers for 1 month or bronchitis, pneumonitis, or oesophagitis (onset after age 1 month)	Definitive	Microscopy, culture, or antigen detection in affected tissue.
Histoplasmosis, disseminated or extrapulmonary	Definitive	Microscopy, culture, or antigen detection in affected tissue.
Isosporiasis, with diarrhoea for over 1 month	Definitive	Microscopy (histology or cytology).
Kaposi's sarcoma	Definitive *Presumptive*	Microscopy (histology or cytology). *Characteristic erythematous/violaceous plaque-like lesion on skin or mucous membrane.*
Lymphoma: Burkitt's or immunoblastic or primary in the brain	Definitive	Microscopy (histology or cytology).
Mycobacteriosis: disseminated (including extrapulmonary TB)	Definitive *Presumptive*	Culture. *AFB (species not identified by culture) on microscopy of stool specimen or normally sterile body fluid/tissue, not lungs, skin, cervical or hilar nodes.*
*Mycobacteriosis: pulmonary tuberculosis	Definitive *Presumptive*	Culture or other definitive demonstation of M.*tuberculosis* infection. *Clinical diagnosis, with or without AFB on microscopy, resulting in initiation of anti TB therapy.*
Pneumocystis carinii pneumonia	Definitive *Presumptive*	Microscopy (histology or cytology). *Recent onset dyspnoea on exertion or dry cough and diffuse bilateral pneumonia.*
*Pneumonia recurrent within a 12 month period	Definitive *Presumptive*	Two episodes proven microbiologically. *CXR or clinical diagnosis.*
Progressive multifocal leukoencephalopathy (PML)	Definitive	Electron Microscopy, antigen detection in brain and urine, antibody in serum or CSF.
Salmonella (non-typhoid) septicaemia recurrent	Definitive	Culture.
Toxoplasmosis of brain (onset after age 1 month)	Definitive *Presumptive*	Microscopy (histology or cytology), mouse inoculation, tissue culture. *Recent onset focal neurological abnormality or reduced level of consciousness, and mass effect lesion on scan, and serological evidence or specify therapy response.*
Wasting syndrome due to HIV	Definitive	Weight loss (over 10% baseline) with no other cause, and 30 days or more of either diarrhoaea or weakness with fever.
**Bacterial infections, multiple or recurrent in a child aged less than 13 years	Definitive	Culture, antigen detection, CSF microscopy.
**Lymphoid interstitial pneumonia and/or pulmonary lymphoid hyperplasia in a child less than 13 years	Definitive *Presumptive*	Microscopy (histology or cytology). *Diffuse bilateral reticulonodular pulmonary interstitial infiltrates for over 2 months and no pathogen identified and no antibiotic response.*

* Applicable only to patients aged ≥ 13 years diagnosed with these conditions on or after 1 January 1993.
** Applicable only to patients aged < 13 years when diagnosed.

Appendix C: Supplementary data on evaluating algorithms to determine estART

C.1 Evaluating a smoothed increase in CD4 counts

After the smoothing mechanism was applied, the evaluation of the algorithm included 8,945 (82.1%) of the random sample of 10,897 individuals (Figure C.1). There was a peak in the likelihood ratios and the cut-off of 110 cells/mm^3 had the highest likelihood ratio (highlighted in Table C.1) and was selected to produce the results.

Figure C.1. Distribution of CD4 counts in relation to startART

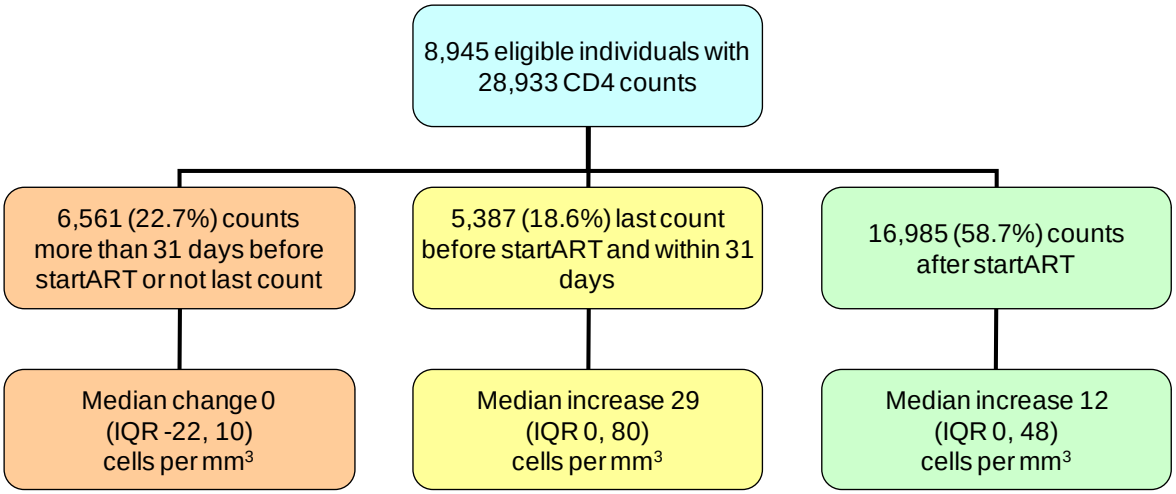


Table C.1. Summary table for a single increase in smoothed CD4 counts

Cut-off for change in CD4 counts / cells/mm^3	Number of positives	Number of true positives	Number of negatives	Number of true negatives	Sensitivity	Specificity	Correctly Classified	Likelihood ratio
0	8,650	3,903	295	168	61%	7%	46%	0.655
20	6,866	3,811	2,079	944	60%	37%	53%	0.945
40	5,595	3,170	3,350	1,343	50%	52%	50%	1.041
60	4,344	2,458	4,601	1,672	39%	65%	46%	1.102
80	3,282	1,886	5,663	1,921	30%	75%	43%	1.168
100	2,408	1,404	6,537	2,127	22%	83%	39%	1.271
110	2,013	1,190	6,932	2,212	19%	86%	38%	1.331
120	1,472	1,028	7,473	2,261	16%	88%	37%	1.330
130	1,492	885	7,453	2,304	14%	90%	36%	1.328
140	1,284	756	7,661	2,336	12%	91%	35%	1.288
160	930	544	8,015	2,390	8.5%	93%	33%	1.200
180	657	386	8,288	2,445	6.1%	95%	32%	1.218
200	482	282	8,463	2,478	4.4%	96%	31%	1.199

C.2 Evaluating a preceding decrease and subsequent increase in CD4 counts

The evaluation of the algorithm included 9,284 (85.2%) of the random sample of 10,897 individuals (Figure C.2). Of their 30,401 CD4 cell counts, 8,479 (27.9%) were preceded by a decrease and followed by an increase in CD4 counts. The remaining 21,922 CD4 counts were not preceded by a decrease and followed by an increase in CD4 counts: 9,688 were preceded and followed by increases, 7,937 were preceded by an increase and followed by a decrease, 4,206 were preceded and followed by decreases, and 91 were preceded and followed by no change.

With this algorithm the likelihood ratios changed little using the different cut-offs. The cut-off of 150 cells/mm^3 (highlighted in Table C.2) was selected to produce the results.

Figure C.2. Distribution of CD4 counts in relation to startART

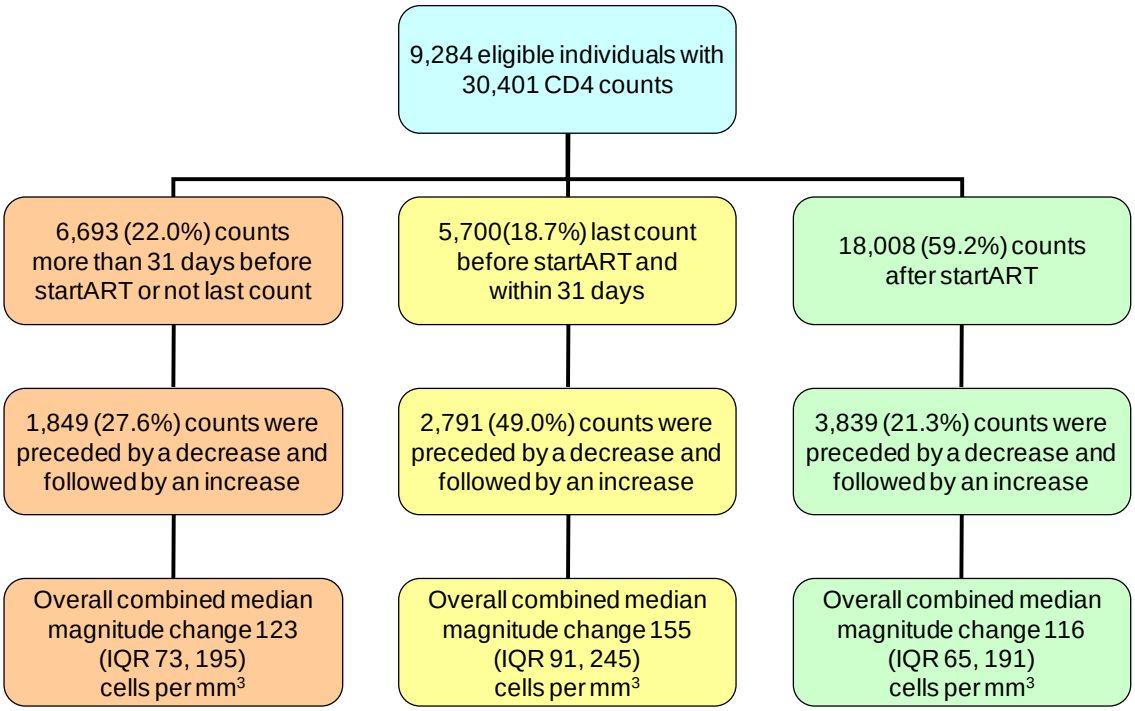


Table C.2. Summary table for a preceding decrease and subsequent increase in CD4 counts

Cut-off for change in CD4 counts / cells/mm^3	Number of positives	Number of true positives	Number of negatives	Number of true negatives	Sensitivity	Specificity	Correctly Classified	Likelihood ratio
0	6,249	2,706	3,035	1,320	40%	51%	43%	0.824
10	6,227	2,696	3,057	1,323	40%	51%	43%	0.822
20	6,140	2,660	3,144	1,345	40%	52%	43%	0.826
30	6,015	2,590	3,269	1,376	39%	53%	43%	0.825
40	5,826	2,509	3,458	1,416	38%	55%	42%	0.826
50	5,593	2,419	3,691	1,473	36%	57%	42%	0.837
60	5,345	2,310	3,939	1,527	35%	59%	41%	0.839
70	5,099	2,204	4,185	1,571	33%	61%	41%	0.835
80	4,873	2,121	4,411	1,616	32%	62%	40%	0.841
90	4,612	2,039	4,672	1,678	30%	65%	40%	0.863
100	4,349	1,934	4,935	1,738	29%	67%	40%	0.876
110	4,077	1,825	5,207	1,799	27%	69%	39%	0.890
120	3,824	1,724	5,460	1,853	26%	71%	39%	0.902
130	3,572	1,607	5,712	1,905	24%	73%	38%	0.904
140	3,326	1,510	5,958	1,943	23%	75%	37%	0.899
150	3,089	1,409	6,195	1,994	21%	77%	37%	0.911
160	2,872	1,304	6,412	2,028	19%	78%	36%	0.893
170	2,684	1,223	6,600	2,057	18%	79%	35%	0.883
180	2,471	1,134	6,813	2,102	17%	81%	35%	0.894
190	2,274	1,038	7,010	2,146	16%	83%	34%	0.898
200	2,106	963	7,178	2,177	14%	84%	34%	0.895
220	1,942	886	7,342	2,204	13%	85%	33%	0.881
240	1,788	812	7,496	2,227	12%	86%	33%	0.858
260	1,665	754	7,619	2,251	11%	87%	32%	0.852
280	1,532	684	7,752	2,273	10.2%	88%	32%	0.826
300	1,413	636	7,871	2,294	9.5%	88%	32%	0.822

C.3 Evaluating a preceding decrease and subsequent increase in smoothed CD4 counts

The evaluation of the algorithm included 7,513 (68.9%) of the random sample of 10,897 individuals (Figure C.3). There were 5,063 CD4 counts that were preceded by a decrease or no change and followed by an increase in smoothed CD4 counts. The remaining 18,436 CD4 counts were not preceded by a decrease and followed by an increase in smoothed CD4 counts: 7,123 were preceded and followed by increases, 5,230 were preceded by an increase and followed by a decrease, 4,715 were preceded and followed by decreases, and 1,368 were preceded and followed by no change.

There was a clear but minor peak in the likelihood ratios with this algorithm and the cut-off of 130 cells/mm^3 had the highest likelihood ratio (highlighted in Table C.3) and was selected to produce the results.

Figure C.3. Distribution of CD4 counts in relation to startART

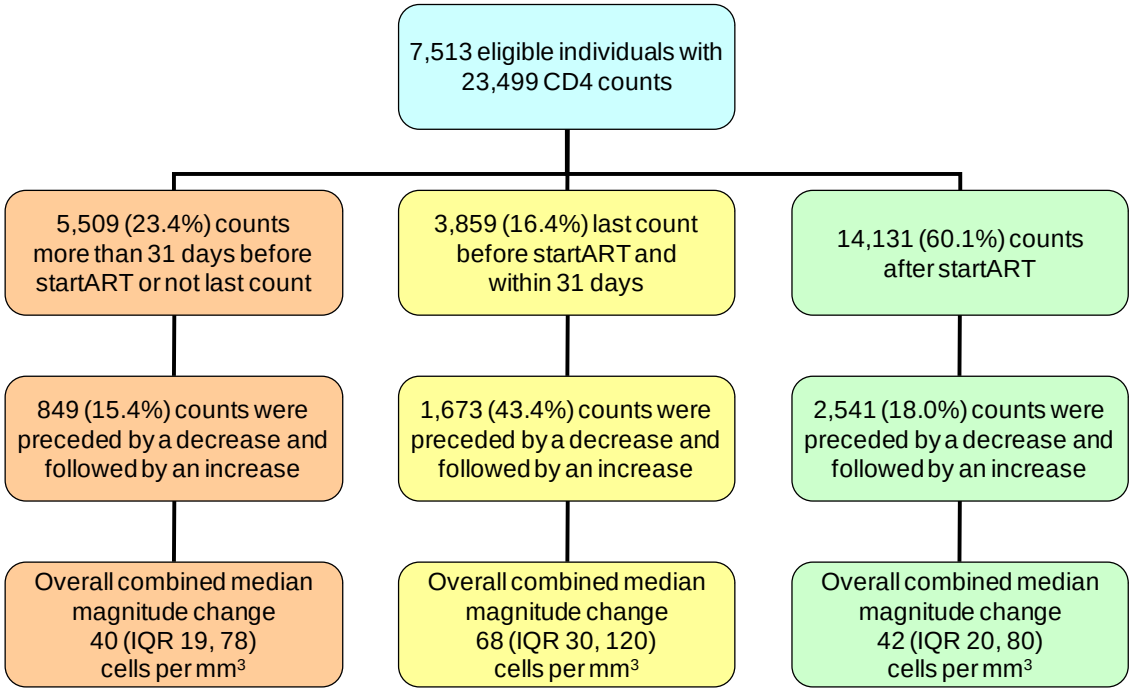


Table C.3. Summary table for a preceding decrease and subsequent increase in smoothed CD4 counts

Cut-off for change in CD4 counts / cells/mm^3	Number of positives	Number of true positives	Number of negatives	Number of true negatives	Sensitivity	Specificity	Correctly Classified	Likelihood ratio
0	4,320	2,030	3,193	1,950	45%	64%	53%	1.27
10	4,006	1,902	3,507	2,053	42%	68%	53%	1.31
20	3,561	1,732	3,952	2,176	39%	72%	52%	1.37
30	3,145	1,571	4,368	2,305	35%	76%	52%	1.46
40	2,745	1,398	4,768	2,408	31%	79%	51%	1.51
50	2,362	1,225	5,151	2,511	27%	83%	50%	1.58
60	2,065	1,082	5,448	2,599	24%	86%	49%	1.68
70	1,777	943	5,736	2,652	21%	87%	48%	1.67
80	1,538	817	5,975	2,701	18%	89%	47%	1.66
90	1,315	720	6,198	2,763	16%	91%	46%	1.79
100	1,154	620	6,359	2,790	14%	92%	45%	1.72
110	962	526	6,551	2,837	11.7%	93%	45%	1.80
120	829	456	6,684	2,864	10.2%	94%	44%	1.81
130	707	394	6,806	2,890	8.8%	95%	44%	1.84
140	607	335	6,906	2,905	7.5%	96%	43%	1.75
150	525	288	6,988	2,925	6.4%	96%	43%	1.77

C.4 Evaluating a positive slope in CD4 counts

The evaluation of the algorithm included 9,284 (85.2%) of the random sample of 10,897 individuals (Figure C.4). Using this algorithm the specificity approached 100% as the cut-offs increased and there was no peak in the likelihood ratios. Therefore, the cut-off of 800 cells/mm^3 per year (highlighted in Table C.4) was selected as it resulted in a similar sensitivity (22.5%) as the previous algorithms.

Figure C.4. Distribution of CD4 counts in relation to startART

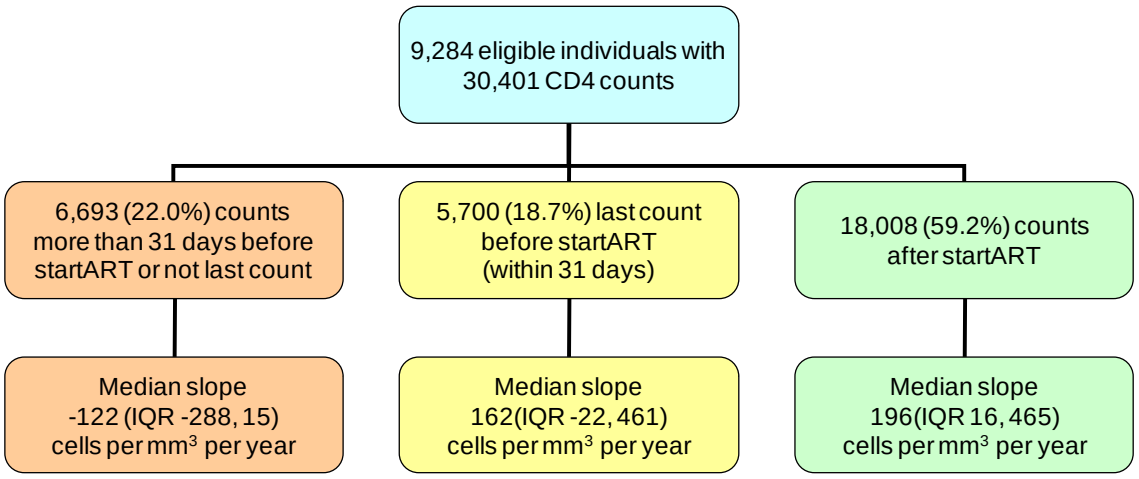


Table C.4. Summary table for a slope in CD4 counts

Cut-off for change in CD4 counts / cells/mm^3 per year	Number of positives	Number of true positives	Number of negatives	Number of true negatives	Sensitivity	Specificity	Correctly Classified	Likelihood ratio
100	7,880	4,486	1,404	640	67%	25%	55%	0.89
200	6,791	4,023	2,493	1,109	60%	43%	55%	1.05
300	5,635	3,462	3,649	1,531	52%	59%	54%	1.26
400	4,659	2,944	4,625	1,824	44%	70%	51%	1.48
500	3,837	2,493	5,447	2,061	37%	79%	49%	1.81
600	3,171	2,109	6,113	2,182	32%	84%	46%	1.98
700	2,651	1,796	6,633	2,276	27%	88%	44%	2.19
800	2,209	1,508	7,075	2,351	23%	91%	42%	2.41
900	1,844	1,283	7,440	2,410	19%	93%	40%	2.70
1,000	1,569	1,088	7,715	2,445	16%	94%	38%	2.83
1,100	1,337	927	7,947	2,472	14%	95%	37%	2.95
1,200	1,130	786	8,154	2,495	12%	96%	35%	3.08
1,300	977	690	8,307	2,512	10%	97%	34%	3.26
1,400	850	597	8,434	2,518	9%	97%	34%	3.05
1,500	728	508	8,556	2,533	8%	98%	33%	3.23
1,600	629	446	8,655	2,546	6.7%	98%	32%	3.60
1,700	552	393	8,732	2,552	5.9%	98%	32%	3.63
1,800	489	354	8,795	2,559	5.3%	99%	31%	3.92

C.5 Evaluating a preceding negative slope and subsequent positive slope in CD4 counts

The evaluation included 7,513 (68.9%) of the sample of 10,897 individuals (Figure C.5). There were 7,448 counts that were preceded by a negative slope and followed by a positive slope. Of the remaining 16,051 counts: 8,223 were preceded and followed by positive slopes, 4,845 were preceded by a positive slope and followed by a negative slope, 2,979 were preceded and followed by negative slopes, and 4 were preceded and followed by slopes of zero.

Using this algorithm the specificity approached 100% as the cut-offs increased and there was no peak in the likelihood ratios. Therefore, the cut-off of 800 cells/mm^3 per year (highlighted in Table C.5) was selected as it resulted in a similar sensitivity (22.4%) as the previous algorithms.

Figure C.5. Distribution of CD4 counts in relation to startART

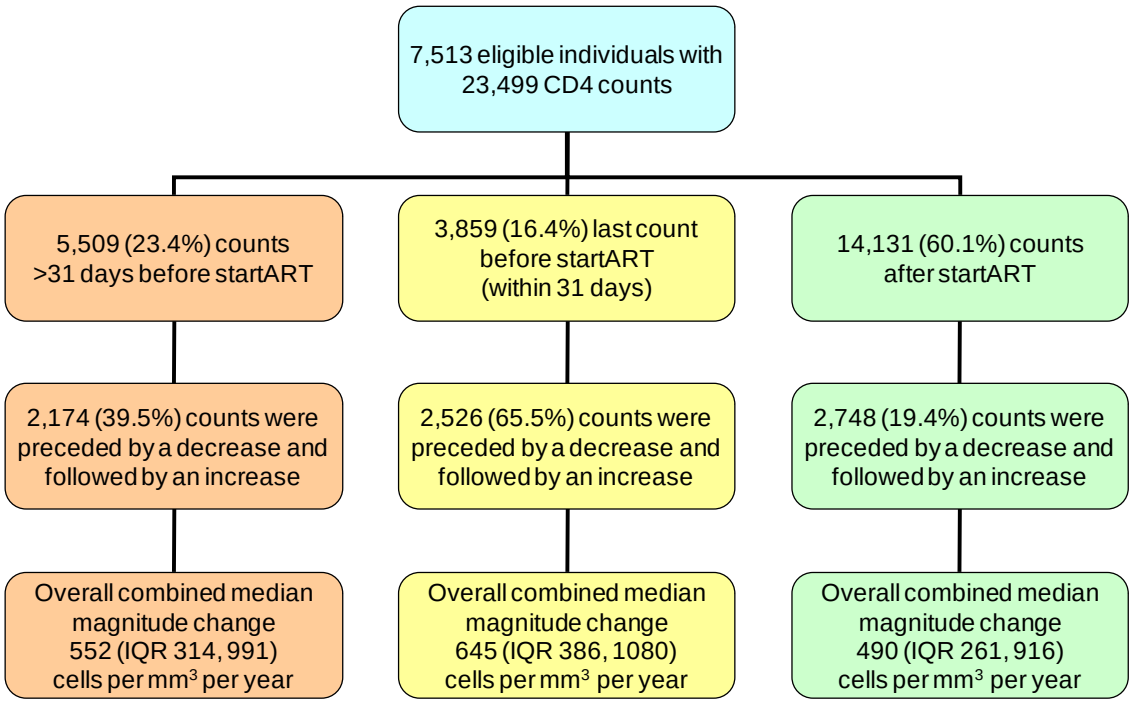


Table C.5. Summary table for a preceding negative slope and subsequent positive slope in CD4 counts

Cut-off for change in CD4 counts / cells/mm^3 per year	Number of positives	Number of true positives	Number of negatives	Number of true negatives	Sensitivity	Specificity	Correctly Classified	Likelihood ratio
100	4,459	1,637	3,054	1,947	37%	64%	48%	1.02
200	4,146	1,591	3,367	2,080	36%	69%	49%	1.13
300	3,741	1,522	3,772	2,243	34%	74%	50%	1.30
400	3,331	1,456	4,182	2,382	33%	78%	51%	1.51
500	2,897	1,365	4,616	2,529	30%	83%	52%	1.83
600	2,518	1,255	4,995	2,634	28%	87%	52%	2.12
700	2,143	1,091	5,370	2,724	24%	90%	51%	2.38
800	1,881	1,001	5,632	2,783	22%	92%	50%	2.69
900	1,653	905	5,860	2,834	20%	93%	50%	3.05
1,000	1,440	814	6,073	2,867	18%	94%	49%	3.28
1,100	1,243	720	6,270	2,900	16%	96%	48%	3.61
1,200	1,071	640	6,442	2,920	14%	96%	47%	3.77
1,300	929	577	6,584	2,943	13%	97%	47%	4.25
1,400	817	513	6,696	2,960	11%	98%	46%	4.64
1,500	716	459	6,797	2,971	10%	98%	46%	4.86
1,600	627	399	6,886	2,978	8.9%	98%	45%	4.74

References

(1) Du Bois RM, Branthwaite MA, Mikhail JR, Batten JC. Primary pneumocystis carinii and cytomegalovirus infection. *Lancet* 1981; ii:1339.

(2) Gill ON, Adler MW, Day NE. Monitoring the prevalence of HIV. *BMJ* 1989; 299(6711):1295-1298.

(3) Department of Health. The National Health Service Act 1977: The NHS Trusts and Primary Care Trusts (Sexually Transmitted Diseases) Directions 2000. 8-9-2000. Crown.

(4) PHLS Communicable Disease Surveillance Centre. Quarterly Communicable Disease Review October to December 2001: Caldicott, Confidentiality and the Patient Information Advisory Group. *J Public Health Med* 2002; 24:138-144.

(5) Barré-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* 1983; 220:868-871.

(6) Sarngadharan MG, Popovic M, Bruch L, Schupbach J, Gallo RC. Antibodies reactive with human T-lymphotropic retroviruses (HTLV-III) in the serum of patients with AIDS. *Science* 1984; 224(4648):506-508.

(7) Gallo RC, Sarin PS, Gelmann EP, Robert-Guroff M, Richardson E, Kalyanaraman VS et al. Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science* 1983; 224:500-503.

(8) Gallo RC, Salahuddin SZ, Popovic M, Shearer GM, Kaplan M, Haynes BF et al. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science* 1984; 224(4648):500-503.

(9) Popovic M, Sarngadharan MG, Read E, Gallo RC. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. *Science* 1984; 224(4648):497-500.

(10) Schupbach J, Popovic M, Gilden RV, Gonda MA, Sarngadharan MG, Gallo RC. Serological analysis of a subgroup of human T-lymphotropic retroviruses (HTLV-III) associated with AIDS. *Science* 1984; 224(4648):503-505.

(11) Centers for Disease Control. Pneumocystis pneumonia - Los Angeles. *MMWR* 1981; 30:250-252.

(12) Dybul M, Fauci AS, Bartlett JG, Kaplan JE, Pau AK. Guidelines for using antiretroviral agents among HIV-infected adults and adolescents. *Ann Intern Med* 2002; 137(5 Pt 2):381-433.

(13) European Centre for the Epidemiological Monitoring of AIDS. 1993 revision of the European AIDS surveillance case definition. AIDS Surveillance in Europe, Quarterly Report 1993. 37, 23-28. 1993.

(14) Ancelle-Park R. Expanded European AIDS case definition. *Lancet* 1993; 341(8842):441.

(15) Dalgleish AG, Beverley PC, Clapham PR, Crawford DH, Greaves MF, Weiss RA. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature* 1984; 312(5996):763-767.

(16) Klatzmann D, Barre-Sinoussi F, Nugeyre MT, Danquet C, Vilmer E, Griscelli C et al. Selective tropism of lymphadenopathy associated virus (LAV) for helper-inducer T lymphocytes. *Science* 1984; 225(4657):59-63.

(17) Klatzmann D, Champagne E, Chamaret S, Gruest J, Guetard D, Hercend T et al. T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. *Nature* 1984; 312(5996):767-768.

(18) Gottlieb MS, Schroff R, Schanker HM, Weisman JD, Thim Fan P, Wolf RA et al. Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men. *N Engl J Med* 1981; 305:1425-1431.

(19) Mathur-Wagh U, Enlow RW, Spigland I, Winchester RJ, Sacks HS, Rorat E et al. Longitudinal study of persistent generalised lymphadenopathy in homosexual men: relation to acquired immunodeficiency syndrome. *Lancet* 1984; 1(8385):1033-1038.

(20) Schwartz K, Visscher BR, Detels R, Taylor J, Nishanian P, Fahey JL. Immunological changes in lymphadenopathy virus positive and negative symptomless male homosexuals: two years of observation. *Lancet* 1985; 2(8459):831-832.

(21) Melbye M, Biggar RJ, Ebbesen P, Neuland C, Goedert JJ, Faber V et al. Long-term seropositivity for human T-lymphotropic virus type III in homosexual men without the acquired immunodeficiency syndrome: development of immunologic and clinical abnormalities. A longitudinal study. *Annals of Internal Medicine* 1986; 104(4):496-500.

(22) Ammann AJ, Abrams D, Conant M, Chudwin D, Cowan M, Volberding P et al. Acquired immune dysfunction in homosexual men: immunologic profiles. *Clinical Immunology & Immunopathology* 1983; 27(3):315-325.

(23) Fahey JL, Prince H, Weaver M, Groopman J, Visscher B, Schwartz K et al. Quantitative changes in T helper or T suppressor/cytotoxic lymphocyte subsets that distinguish acquired immune deficiency syndrome from other immune subset disorders. *American Journal of Medicine* 1984; 76(1):95-100.

(24) Eyster ME, Gail MH, Ballard JO, Al Mondhiry H, Goedert JJ. Natural history of human immunodeficiency virus infections in hemophiliacs: effects of T-cell subsets, platelet counts, and age. *Annals of Internal Medicine* 1987; 107(1):1-6.

(25) Fahey JL, Giorgi J, Martinez-Maza O, Detels R, Mitsuyasu R, Taylor J. Immune pathogenesis of AIDS and related syndromes. *Ann Inst Pasteur Immunol* 1987; 138(2):245-252.

(26) Gerstoft J, Petersen CS, Kroon S, Ullman S, Lindhardt BO, Hofmann B et al. The immunological and clinical outcome of HIV infection: 31 months of follow-up in a cohort of homosexual men. *Scand J Infect Dis* 1987; 19(5):503-509.

(27) Polk BF, Fox R, Brookmeyer R, Kanchanaraksa S, Kaslow R, Visscher B et al. Predictors of the acquired immunodeficiency syndrome developing in a cohort of seropositive homosexual men. *New England Journal of Medicine* 1987; 316(2):61-66.

(28) Goedert JJ, Biggar RJ, Melbye M, Mann DL, Wilson S, Gail MH et al. Effect of T4 count and cofactors on the incidence of AIDS in homosexual men infected with human immunodeficiency virus. *JAMA* 1987; 257(3):331-334.

(29) Cozzi Lepri A, Sabin CA, Pezzotti P, England PD, Phillips AN, Rezza G. Is there a general tendency for CD4 lymphocyte decline to speed up during human immunodeficiency virus infection? Evidence from the Italian Seroconversion Study. *J Infect Dis* 1997; 175(4):775-780.

(30) Galai N, Munoz A, Chen K, Carey VJ, Chmiel J, Zhou SY. Tracking of markers and onset of disease among HIV-1 seroconverters. *Statistics in Medicine* 1993; 12(22):2133-2145.

(31) Hughes MD, Stein DS, Gundacker HM, Valentine FT, Phair JP, Volberding PA. Within-Subject Variation in CD4 Lymphocyte Count in Assymptomatic Human Immunodeficiency Virus Infection: Implications for Patient Monitoring. *J Infect Dis* 1994; 169:28-36.

(32) Longini IM, Jr., Clark WS, Gardner LI, Brundage JF. The dynamics of CD4+ T-lymphocyte decline in HIV-infected individuals: a Markov

modeling approach. *Journal of Acquired Immune Deficiency Syndromes* 1991; 4(11):1141-1147.

(33) Zeger SL, Diggle PJ. Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics* 1994; 50(3):689-699.

(34) Phillips AN, Lee CA, Elford J, Janossy G, Kernoff PB. The cumulative risk of AIDS as the CD4 lymphocyte count declines. *Journal of Acquired Immune Deficiency Syndromes* 1992; 5(2):148-152.

(35) Phillips AN, Sabin CA, Elford J, Bofill M, Emery V, Griffiths PD et al. Viral burden in HIV infection. *Nature* 1994; 367(6459):124.

(36) Ye P, Kirschner DE, Kourtis AP. The thymus during HIV disease: role in pathogenesis and in immune recovery. *Curr HIV Res* 2004; 2(2):177-183.

(37) Meissner EG, Duus KM, Loomis R, D'Agostin R, Su L. HIV-1 replication and pathogenesis in the human thymus. *Curr HIV Res* 2003; 1(3):275-285.

(38) Phillips AN, Lee CA, Elford J, Webster A, Janossy G, Timms A et al. More rapid progression to AIDS in older HIV-infected people: the role of CD4+ T-cell counts. *Journal of Acquired Immune Deficiency Syndromes* 1991; 4(10):970-975.

(39) Viard JP, Mocroft A, Chiesi A, Kirk O, Roge B, Panos G et al. Influence of age on CD4 cell recovery in human immunodeficiency virus-infected patients receiving highly active antiretroviral therapy: evidence from the EuroSIDA study. *J Infect Dis* 2001; 183(8):1290-1294.

(40) Touloumi G, Karafoulidou A, Gialeraki A, Katsarou O, Milona I, Kapsimali V et al. Determinants of progression of HIV infection in a Greek hemophilia cohort followed for up to 16 years after seroconversion. *Journal of acquired immune deficiency syndromes and human retrovirology* 1998; 19(1):89-97.

(41) Babiker AG, Peto T, Porter K, Walker AS, Darbyshire JH. Age as a determinant of survival in HIV infection. *J Clin Epidemiol* 2001; 54 Suppl 1:S16-S21.

(42) Bofill M, Janossy G, Lee CA, Macdonald-Burns D, Phillips AN, Sabin CA et al. Laboratory control values for CD4 and CD8 T lymphocytes. Implications for HIV-1 diagnosis. *Clin Exp Immunol* 1992; 88:243-252.

(43) Maini MK, Gilson RJ, Chavda N, Gill S, Fakoya A, Ross EJ et al. Reference ranges and sources of variability of CD4 counts in HIV-seronegative women and men. *Genitourinary Medicine* 1996; 72(1):27-31.

(44) Mair C, Hawes SE, Agne HD, Sow PS, N'doye I, Manhart LE et al. Factors associated with CD4 lymphocyte counts in HIV-negative Senegalese individuals. *Clin Exp Immunol* 2008; 151(3):432-440.

(45) Reichert T, DeBruyere M, Deneys V, Totterman T, Lydyard P, Yuksel F et al. Lymphocyte subset reference ranges in adult Caucasians. *Clinical Immunology & Immunopathology* 1991; 60(2):190-208.

(46) Giorgi JV, Cheng HL, Margolick JB, Bauer KD, Ferbas J, Waxdal M et al. Quality control in the flow cytometric measurement of T-lymphocyte subsets: the multicenter AIDS cohort study experience. The Multicenter AIDS Cohort Study Group. *Clinical Immunology & Immunopathology* 1990; 55(2):173-186.

(47) Williams BG, Korenromp EL, Gouws E, Schmid GP, Auvert B, Dye C. HIV infection, antiretroviral therapy, and CD4+ cell count distributions in African populations. *J Infect Dis* 2006; 194(10):1450-1458.

(48) Malone JL, Simms TE, Gray GC, Wagner KF, Burge JR, Burke DS. Sources of Variability in Repeated T- Helper Lymphocyte Counts from Human Immunodeficiency Virus Type 1-Infected Patients: Total Lymphocyte Count Fluctuations and Diurnal Cycle Are Important. *J Acquir Immune Defic Syndr* 1990; 3:144-151.

(49) Prins M, Meyer L, Hessol NA. Sex and the course of HIV infection in the pre- and highly active antiretroviral therapy eras. *AIDS* 2005; 19(4):357-370.

(50) Chama CM, Morrupa JY, Abja UA, Kayode A. Normal CD4 T-lymphocyte baseline in healthy HIV-negative pregnant women. *J Obstet Gynaecol* 2009; 29(8):702-704.

(51) Torres AJ, Angelo AL, Netto EM, Sampaio GP, Souza DF, Inocencio LA et al. Reference range for T lymphocytes populations in blood donors from two different regions in Brazil. *Braz J Infect Dis* 2009; 13(3):221-225.

(52) Murugavel KG, Balakrishnan P, Mohanakrishnan J, Solomon SS, Shankar EM, Muthu Sundaram SP et al. Establishment of T-lymphocyte subset reference intervals in a healthy adult population in Chennai, India. *Indian J Med Res* 2009; 129(1):59-63.

(53) Das BR, Bhanushali AA, Khadapkar R, Jeswani KD, Bhavsar M, Dasgupta A. Reference ranges for lymphocyte subsets in adults from western India: influence of sex, age and method of enumeration. *Indian J Med Sci* 2008; 62(10):397-406.

(54) Prins M, Robertson JR, Brettle RP, Aguado IH, Broers B, Boufassa F et al. Do gender differences in CD4 cell counts matter? *AIDS* 1999; 13(17):2361-2364.

(55) Delmas MC, Jadand C, De V, I, Deveau C, Persoz A, Sobel A et al. Gender difference in CD4+ cell counts persist after HIV-1 infection. SEROCO Study Group. *AIDS* 1997; 11(8):1071-1073.

(56) CASCADE (Concerted Action on SeroConversion to AIDS and Death in Europe) Collaboration. Differences in CD4 cell counts at seroconversion and decline among 5739 HIV-1-infected individuals with well-estimated dates of seroconversion. *J Acquir Immune Defic Syndr* 2003; 34(1):76-83.

(57) Towers CV, Rumney PJ, Ghamsary MG. Longitudinal study of CD4+ cell counts in HIV-negative pregnant patients. *J Matern Fetal Neonatal Med* 2010; 23(10):1091-1096.

(58) Tsegaye A, Wolday D, Otto S, Petros B, Assefa T, Alebachew T et al. Immunophenotyping of blood lymphocytes at birth, during childhood, and during adulthood in HIV-1-uninfected Ethiopians. *Clin Immunol* 2003; 109(3):338-346.

(59) Kibaya RS, Bautista CT, Sawe FK, Shaffer DN, Sateren WB, Scott PT et al. Reference ranges for the clinical laboratory derived from a rural population in Kericho, Kenya. *PLoS ONE* 2008; 3(10):e3327.

(60) Jiang W, Kang L, Lu HZ, Pan X, Lin Q, Pan Q et al. Normal values for CD4 and CD8 lymphocyte subsets in healthy Chinese adults from Shanghai. *Clin Diagn Lab Immunol* 2004; 11(4):811-813.

(61) Webster HK, Pattanapanyasat K, Phanupak P, Wasi C, Chuenchitra C, Ybarra L et al. Lymphocyte immunophenotype reference ranges in healthy Thai adults: implications for management of HIV/AIDS in Thailand. *Southeast Asian J Trop Med Public Health* 1996; 27(3):418-429.

(62) Lee BW, Yap HK, Chew FT, Quah TC, Prabhakaran K, Chan GS et al. Age- and sex-related changes in lymphocyte subpopulations of healthy Asian subjects: from birth to adulthood. *Cytometry* 1996; 26(1):8-15.

(63) Mekonnen Y, Geskus RB, Hendriks JC, Messele T, Borghans J, Miedema F et al. Low CD4 T cell counts before HIV-1 seroconversion do not affect disease progression in Ethiopian factory workers. *J Infect Dis* 2005; 192(5):739-748.

(64) Al Qouzi A, Al Salamah A, Al Rasheed R, Al Musalam A, Al Khairy K, Kheir O et al. Immunophenotyping of peripheral blood lymphocytes in Saudi men. *Clin Diagn Lab Immunol* 2002; 9(2):279-281.

(65) Uppal SS, Verma S, Dhot PS. Normal values of CD4 and CD8 lymphocyte subsets in healthy indian adults and the effects of sex, age, ethnicity, and smoking. *Cytometry B Clin Cytom* 2003; 52(1):32-36.

(66) Mandala WL, MacLennan JM, Gondwe EN, Ward SA, Molyneux ME, Maclennan CA. Lymphocyte subsets in healthy Malawians: implications for immunologic assessment of HIV infection in Africa. *J Allergy Clin Immunol* 2010; 125(1):203-208.

(67) Lugada ES, Mermin J, Kaharuza F, Ulvestad E, Were W, Langeland N et al. Population-based hematologic and immunologic reference values for a healthy Ugandan population. *Clin Diagn Lab Immunol* 2004; 11(1):29-34.

(68) Levin A, Brubaker G, Shao JS, Kumby D, O'Brien TR, Goedert JJ et al. Determination of T-lymphocyte subsets on site in rural Tanzania: results in HIV-1 infected and non-infected individuals. *Int J STD AIDS* 1996; 7(4):288-291.

(69) Tugume SB, Piwowar EM, Lutalo T, Mugyenyi PN, Grant RM, Mangeni FW et al. Hematological reference ranges among healthy Ugandans. *Clin Diagn Lab Immunol* 1995; 2(2):233-235.

(70) Amatya R, Vajpayee M, Kaushik S, Kanswal S, Pandey RM, Seth P. Lymphocyte immunophenotype reference ranges in healthy Indian adults: implications for management of HIV/AIDS in India. *Clin Immunol* 2004; 112(3):290-295.

(71) Chng WJ, Tan GB, Kuperan P. Establishment of adult peripheral blood lymphocyte subset reference range for an Asian population by single-platform flow cytometry: influence of age, sex, and race and comparison with other published studies. *Clin Diagn Lab Immunol* 2004; 11(1):168-173.

(72) Howard RR, Fasano CS, Frey L, Miller CH. Reference intervals of CD3, CD4, CD8, CD4/CD8, and absolute CD4 values in Asian and Non-Asian populations. *Cytometry* 1996; 26(3):231-232.

(73) Dhaliwal JS, Balasubramaniam T, Quek CK, Gill HK, Nasuruddin BA. Reference ranges for lymphocyte subsets in a defined Malaysian population. *Singapore Med J* 1995; 36(3):288-291.

(74) Choong ML, Ton SH, Cheong SK. Influence of race, age and sex on the lymphocyte subsets in peripheral blood of healthy Malaysian adults. *Ann Clin Biochem* 1995; 32 (Pt 6):532-539.

(75) Prince HE, Hirji K, Waldbeser LS, Plaeger-Marshall S, Kleinman S, Lanier LL. Influence of racial background on the distribution of T-cell subsets and Leu 11-positive lymphocytes in healthy blood donors. *Diagn Immunol* 1985; 3(1):33-37.

(76) Tollerud DJ, Clark JW, Brown LM, Neuland CY, Pankiw-Trost LK, Blattner WA et al. The influence of age, race, and gender on peripheral

blood mononuclear-cell subsets in healthy nonsmokers. *J Clin Immunol* 1989; 9(3):214-222.

(77) Nowicki MJ, Karim R, Mack WJ, Minkoff H, Anastos K, Cohen M et al. Correlates of CD4+ and CD8+ lymphocyte counts in high-risk immunodeficiency virus (HIV)-seronegative women enrolled in the women's interagency HIV study (WIHS). *Hum Immunol* 2007; 68(5):342-349.

(78) Muller V, von W, V, Yerly S, Boni J, Klimkait T, Burgisser P et al. African descent is associated with slower CD4 cell count decline in treatment-naive patients of the Swiss HIV Cohort Study. *AIDS* 2009; 23(10):1269-1276.

(79) Todd J, Glynn JR, Marston M, Lutalo T, Biraro S, Mwita W et al. Time from HIV seroconversion to death: a collaborative analysis of eight studies in six low and middle-income countries before highly active antiretroviral therapy. *AIDS* 2007; 21 Suppl 6:S55-S63.

(80) Del Amo J, Petruckevitch A, Phillips A, Johnson AM, Stephenson J, Desmond N et al. Disease progression and survival in HIV-1-infected Africans in London. *AIDS* 1998; 12(10):1203-1209.

(81) Silverberg MJ, Wegner SA, Milazzo MJ, McKaig RG, Williams CF, Agan BK et al. Effectiveness of highly-active antiretroviral therapy by race/ethnicity. *AIDS* 2006; 20(11):1531-1538.

(82) Pantazis N, Morrison C, Amornkul PN, Lewden C, Salata RA, Minga A 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):e32369.

(83) May M, Wood R, Myer L, Taffe P, Rauch A, Battegay M et al. CD4(+) T cell count decreases by ethnicity among untreated patients with HIV infection in South Africa and Switzerland. *J Infect Dis* 2009; 200(11):1729-1735.

(84) Clark SJ, Saag MS, Decker WD, Campbell-Hill S, Roberson JL, Veldkamp PJ et al. High titers of cytopathic virus in plasma of patients with symptomatic primary HIV-1 infection. *N Engl J Med* 1991; 324(14):954-960.

(85) Niu MT, Stein DS, Schnittman SM. Primary human immunodeficiency virus type 1 infection: review of pathogenesis and early treatment intervention in humans and animal retrovirus infections. *J Infect Dis* 1993; 168(6):1490-1501.

(86) Masel J, Arnaout RA, O'Brien TR, Goedert JJ, Lloyd AL. Fluctuations in HIV-1 viral load are correlated to CD4+ T-lymphocyte count during the

natural course of infection. *J Acquir Immune Defic Syndr* 2000; 23(5):375-379.

(87) Touloumi G, Pantazis N, Babiker AG, Walker SA, Katsarou O, Karafoulidou A et al. Differences in HIV RNA levels before the initiation of antiretroviral therapy among 1864 individuals with known HIV-1 seroconversion dates. *AIDS* 2004; 18(12):1697-1705.

(88) Sabin CA, Devereux H, Phillips AN, Hill A, Janossy G, Lee CA et al. Course of viral load throughout HIV-1 infection. *J Acquir Immune Defic Syndr* 2000; 23(2):172-177.

(89) Henrard DR, Phillips JF, Muenz LR, Blattner WA, Wiesner D, Eyster ME et al. Natural history of HIV-1 cell-free viremia. *JAMA* 1995; 274(7):554-558.

(90) Vidal C, Garcia F, Romeu J, Ruiz L, Miro JM, Cruceta A et al. Lack of evidence of a stable viral load set-point in early stage asymptomatic patients with chronic HIV-1 infection. *AIDS* 1998; 12(11):1285-1289.

(91) Mellors JW, Rinaldo CR, Jr., Gupta P, White RM, Todd JA, Kingsley LA. Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science* 1996; 272(5265):1167-1170.

(92) Mellors JW, Munoz A, Giorgi JV, Margolick JB, Tassoni CJ, Gupta P et al. Plasma viral load and CD4+ lymphocytes as prognostic markers of HIV-1 infection. *Annals of Internal Medicine* 1997; 126(12):946-954.

(93) Sabin CA, Devereux H, Phillips AN, Janossy G, Loveday C, Lee CA. Immune markers and viral load after HIV-1 seroconversion as predictors of disease progression in a cohort of haemophilic men. *AIDS* 1998; 12(11):1347-1352.

(94) Whittle H, Morris J, Todd J, Corrah T, Sabally S, Bangali J et al. HIV-2-infected patients survive longer than HIV-1-infected patients. *AIDS* 1994; 8(11):1617-1620.

(95) Marlink R, Kanki P, Thior I, Travers K, Eisen G, Siby T et al. Reduced rate of disease development after HIV-2 infection as compared to HIV-1. *Science* 1994; 265(5178):1587-1590.

(96) Jaffar S, Wilkins A, Ngom PT, Sabally S, Corrah T, Bangali JE et al. Rate of decline of percentage CD4+ cells is faster in HIV-1 than in HIV-2 infection. *J Acquir Immune Defic Syndr Hum Retrovirol* 1997; 16(5):327-332.

(97) Alaeus A, Lidman K, Bjorkman A, Giesecke J, Albert J. Similar rate of disease progression among individuals infected with HIV-1 genetic subtypes A-D. *AIDS* 1999; 13(8):901-907.

(98) Kiwanuka N, Robb M, Laeyendecker O, Kigozi G, Wabwire-Mangen F, Makumbi FE et al. HIV-1 viral subtype differences in the rate of CD4+ T-cell decline among HIV seroincident antiretroviral naive persons in Rakai district, Uganda. *J Acquir Immune Defic Syndr* 2010; 54(2):180-184.

(99) Easterbrook PJ, Smith M, Mullen J, O'Shea S, Chrystie I, de RA et al. Impact of HIV-1 viral subtype on disease progression and response to antiretroviral therapy. *J Int AIDS Soc* 2010; 13:4.

(100) Kiwanuka N, Laeyendecker O, Robb M, Kigozi G, Arroyo M, McCutchan F et al. Effect of human immunodeficiency virus Type 1 (HIV-1) subtype on disease progression in persons from Rakai, Uganda, with incident HIV-1 infection. *J Infect Dis* 2008; 197(5):707-713.

(101) Baeten JM, Chohan B, Lavreys L, Chohan V, McClelland RS, Certain L et al. HIV-1 subtype D infection is associated with faster disease progression than subtype A in spite of similar plasma HIV-1 loads. *J Infect Dis* 2007; 195(8):1177-1180.

(102) Vasan A, Renjifo B, Hertzmark E, Chaplin B, Msamanga G, Essex M et al. Different rates of disease progression of HIV type 1 infection in Tanzania based on infecting subtype. *Clin Infect Dis* 2006; 42(6):843-852.

(103) Kaleebu P, French N, Mahe C, Yirrell D, Watera C, Lyagoba F et al. Effect of human immunodeficiency virus (HIV) type 1 envelope subtypes A and D on disease progression in a large cohort of HIV-1-positive persons in Uganda. *J Infect Dis* 2002; 185(9):1244-1250.

(104) Health Protection Agency CFI. Anti-HIV screening – minimum testing algorithm. National Standard Method VSOP 11(1). 2007.

(105) British HIV Association, British Association for Sexual Health and HIV, British Infection Society. UK National Guidelines for HIV Testing 2008. 1-31. 2008. London, British HIV Association.

(106) Palfreeman A, Fisher M. Diagnosis and management of HIV infection. *Br J Hosp Med (Lond)* 2011; 72(3):146-150.

(107) Palfreeman A, Fisher M, Stewart E. HIV testing in primary care will help improve early diagnosis. *Practitioner* 2009; 253(1723):27-30, 3.

(108) Palfreeman A, Fisher M, Ong E, Wardrope J, Stewart E, Castro-Sanchez E et al. Testing for HIV: concise guidance. *Clin Med* 2009; 9(5):471-476.

(109) Williams I, Ahmed I, Friedland JS, Fisher M, Ong E, Palfreeman A. HIV testing in the UK. *Lancet* 2009; 374(9687):376.

(110) Department of Health. Screening for infectious diseases in pregnancy. London: Crown; 2003.

(111) Department of Health. The national strategy for sexual health and HIV. London: Crown; 2001.

(112) Goubar A, Ades AE, De Angelis D, McGarrigle CA, Mercer C, Tookey PA et al. Bayesian multi-parameter synthesis of HIV surveillance data in England and Wales, 2001. Technical report. 2001. London: Health Protection Agency Centre for Infections.

(113) Goubar A, Ades AE, De Angelis D, McGarrigle C, Mercer C, Tookey PA et al. Estimates of HIV prevalence and proportion diagnosed based on Bayesian multi-parameter synthesis of surveillance data. J.R.Statist.Soc.A.In press, with discussion. 2007.

(114) Health Protection Agency Centre for Infections. Bayesian multi-parameter synthesis of HIV surveillance data in England and Wales, 2001. 2006. London, Health Protection Agency Centre for Infections.

(115) The UK Collaborative Group for HIV and STI Surveillance. Testing Times. HIV and other Sexually Transmitted Infections in the United Kingdom: 2007. London: Health Protection Agency. Centre for Infections.; 2007.

(116) Mocroft A, Katlama C, Johnson AM, Pradier C, Antunes F, Mulcahy F et al. AIDS across Europe, 1994-98: the EuroSIDA study. *Lancet* 2000; 356(9226):291-296.

(117) Macilwain C. 'Better adherence vital in AIDS therapies'. *Nature* 1997; 390(6658):326.

(118) Finzi D, Blankson J, Siliciano JD, Margolick JB, Chadwick K, Pierson T et al. Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. *Nat Med* 1999; 5(5):512-517.

(119) Pierson T, McArthur J, Siliciano RF. Reservoirs for HIV-1: mechanisms for viral persistence in the presence of antiviral immune responses and antiretroviral therapy. *Annu Rev Immunol* 2000; 18:665-708.

(120) Siliciano JD, Kajdas J, Finzi D, Quinn TC, Chadwick K, Margolick JB et al. Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. *Nat Med* 2003; 9(6):727-728.

(121) Hicks CB, Cahn P, Cooper DA, Walmsley SL, Katlama C, Clotet B et al. Durable efficacy of tipranavir-ritonavir in combination with an optimised background regimen of antiretroviral drugs for treatment-experienced HIV-1-infected patients at 48 weeks in the Randomized Evaluation of

Strategic Intervention in multi-drug reSistant patients with Tipranavir (RESIST) studies: an analysis of combined data from two randomised open-label trials. *Lancet* 2006; 368(9534):466-475.

(122) Miller V, Sabin CA, Hertogs K, Bloor S, Martinez-Picado J, D'Aquila R et al. Virological and immunological effects of treatment interruptions in HIV-1 infected patients with treatment failure. *AIDS* 2000; 14(18):2857-2867.

(123) Gazzard BG. British HIV Association Guidelines for the treatment of HIV-1-infected adults with antiretroviral therapy 2008. *HIV Med* 2008; 9(8):563-608.

(124) Egger M, May M, Chene G, Phillips AN, Ledergerber B, Dabis F et al. Prognosis of HIV-1-infected patients starting highly active antiretroviral therapy: a collaborative analysis of prospective studies. *Lancet* 2002; 360(9327):119-129.

(125) Sterling TR, Chaisson RE, Moore RD. HIV-1 RNA, CD4 T-lymphocytes, and clinical response to highly active antiretroviral therapy. *AIDS* 2001; 15(17):2251-2257.

(126) El Sadr WM, Lundgren JD, Neaton JD, Gordin F, Abrams D, Arduino RC et al. CD4+ count-guided interruption of antiretroviral treatment. *N Engl J Med* 2006; 355(22):2283-2296.

(127) National Institute of Allergy and Infectious Diseases. Strategic Timing of AntiRetroviral Treatment (START). 20-3-2009.

(128) Williams I, Churchill D. BHIVA guidelines for the treatment of HIV-1 positive adults with antiretroviral therapy: 2012. British HIV Association [2012 [cited 2012 July 22]; Available from: URL:http://www.bhiva.org/documents/Guidelines/Treatment/2012/120430TreatmentGuidelines.pdf

(129) BHIVA Guidelines Co-ordinating Committee. British HIV Association guidelines for antiretroviral treatment of HIV seropositive individuals. BHIVA Guidelines Co-ordinating Committee. *Lancet* 1997; 349(9058):1086-1092.

(130) Gazzard B, Moyle G. 1998 revision to the British HIV Association guidelines for antiretroviral treatment of HIV seropositive individuals. BHIVA Guidelines Writing Committee. *Lancet* 1998; 352(9124):314-316.

(131) BHIVA Guidelines Co-ordinating Committee. British HIV Association (BHIVA) guidelines for the treatment of HIV-infected adults with antiretroviral therapy. *HIV Med* 2000; 1(2):76-101.

(132) BHIVA Guidelines Co-ordinating Committee. British HIV Association (BHIVA) guidelines for the treatment of HIV-infected adults with antiretroviral therapy. *HIV Med* 2001; 2(4):276-313.

(133) Pozniak A, Gazzard B, Anderson J, Babiker A, Churchill D, Collins S et al. British HIV Association (BHIVA) guidelines for the treatment of HIV-infected adults with antiretroviral therapy. *HIV Med* 2003; 4 Suppl 1:1-41.

(134) Gazzard B. British HIV Association (BHIVA) guidelines for the treatment of HIV-infected adults with antiretroviral therapy (2005). *HIV Med* 2005; 6 Suppl 2:1-61.

(135) Gazzard B, Bernard AJ, Boffito M, Churchill D, Edwards S, Fisher N et al. British HIV Association (BHIVA) guidelines for the treatment of HIV-infected adults with antiretroviral therapy (2006). *HIV Med* 2006; 7(8):487-503.

(136) British HIV Association. Standards for HIV Clinical Care. 2007.

(137) Marx JL. Spread of AIDS sparks new health concern. *Science* 1983; 219(4580):42-43.

(138) Alan M Stanier. How accurate is SOUNDEX Matching? Computers in Genealogy 3[7]. 1990.

(139) Public Health Laboratory Service. The surveillance of Kaposi's sarcoma and opportunistic infections in homosexual males in England and Wales. *Commun Dis Rep* 1982; 34:3-4.

(140) Centers for Disease Control. Update on Kaposi's sarcoma and opportunistic infections in previously healthy persons - United States. *MMWR* 1982; 31:294-301.

(141) Public Health Laboratory Service. Surveillance of the acquired immune deficiency syndrome (AIDS) in the United Kingdom, January 1982 - July 1983. *Commun Dis Rep* 1983; 30:3-5.

(142) Public Health Laboratory Service. Test for antibody to human T-lymphotropic virus III (HTLV III). *Commun Dis Rep* 1984; 84:1.

(143) Evans BG, Gill ON, Emslie JA. Completeness of reporting of AIDS - doctors should beware of 'reporting fatigue'. *BMJ* 1991; 302:1351-1352.

(144) McCormick A. The notification of infectious diseases in England and Wales. *Commun Dis Rep* 1993; 3:R19-R25.

(145) Bacchetti P. Reporting delays of deaths with AIDS in the United States. *J Acquir Immune Defic Syndr Hum Retrovirol* 1996; 13(4):363-367.

(146) Ajdacic-Gross V, Zellweger U, Wang J, Fleerackers Y, Somaini B. How complete is AIDS surveillance in Europe? An eagle eye comparison with mortality data. *J Epidemiol Community Health* 2001; 55(1):52-56.

(147) Modesitt SK, Hulman S, Fleming D. Evaluation of active versus passive AIDS surveillance in Oregon. *Am J Public Health* 1990; 80(4):463-464.

(148) Scheer S, McQuitty M, Denning P, Hormel L, Stephens B, Katz M et al. Undiagnosed and unreported AIDS deaths: results from the San Francisco Medical Examiner. *J Acquir Immune Defic Syndr* 2001; 27(5):467-471.

(149) Hardy AM, Starcher ET, Morgan WM, Druker J, Kristal A, Day JM et al. Review of death certificates to assess completeness of AIDS case reporting. *Public Health Rep* 1987; 102(4):386-391.

(150) Buehler JW, Berkelman RL, Stehr-Green JK. The completeness of AIDS surveillance. *J Acquir Immune Defic Syndr* 1992; 5(3):257-264.

(151) McAnulty JM, Rubin GL, Rushworth RL, Kaldor J. Underreporting of AIDS, New South Wales, 1988-1989. *Med J Aust* 1992; 156(7):452-455.

(152) Buehler JW, Hanson DL, Chu SY. The reporting of HIV/AIDS deaths in women. *Am J Public Health* 1992; 82(11):1500-1505.

(153) Greenberg AE, Hindin R, Nicholas AG, Bryan EL, Thomas PA. The completeness of AIDS case reporting in New York City. *JAMA* 1993; 269(23):2995-3001.

(154) Calzavara LM, Coates RA, Craib KJ, Schechter MT, Le TN, Nault PL et al. Underreporting of AIDS cases in Canada: a record linkage study. *CMAJ* 1990; 142(1):36-39.

(155) Barchielli A, Buiatti E, Galanti C, Giovannetti L, Acciai S, Lazzeri V. Completeness of AIDS reporting and quality of AIDS death certification in Tuscany (Italy): a linkage study between surveillance system of cases and death certificates. *Eur J Epidemiol* 1995; 11(5):513-517.

(156) Evans BG. Estimating underreporting of AIDS: straightforward in theory - difficult in practice. *AIDS* 1991; 5:1261-1262.

(157) Whitmore-Overton SE, Tillett HE, Evans BG, Allardice GM. Improved survival from diagnosis of AIDS in adult cases in the United Kingdom and bias due to reporting delays. *AIDS* 1993; 7(3):415-420.

(158) McCormick A. Trends in mortality statistics in England and Wales with particular reference to AIDS from 1984 to April 1987. *Br Med J (Clin Res Ed)* 1988; 296(6632):1289-1292.

(159) Aldous J, Hickman M, Ellam A, Gazzard B, Hargreaves S. Impact of HIV infection on mortality in young men in a London health authority. *BMJ* 1992; 305(6847):219-221.

(160) King MB. AIDS on the death certificate: the final stigma. *BMJ* 1989; 298(6675):734-736.

(161) Bobby JJ, Spencer PD, Wyatt JC, Farmer RD. AIDS deaths in the UK--how complete are the figures? *Public Health* 1988; 102(6):519-524.

(162) Darby SC, Rizza CR, Doll R, Spooner RJ, Stratton IM, Thakrar B. Incidence of AIDS and excess of mortality associated with HIV in haemophiliacs in the United Kingdom: report on behalf of the directors of haemophilia centres in the United Kingdom. *BMJ* 1989; 298(6680):1064-1068.

(163) McGarrigle C, Gilbart VL, Nicoll A. AIDS and HIV infection acquired heterosexually. Commun Dis Rep Rev 7[9]. 22-9-1997.

(164) Gupta SB, Dingley SD, Lamagni TL, Mortimer JY, Evans BG. The national CD4 surveillance scheme for England and Wales. *Commun Dis Public Health* 2001; 4(1):27-32.

(165) HIV-positive patient retention in the UK: high rate of loss to clinical follow-up among patients from a London clinic. Proceedings of the 17th International AIDS Conference in Mexico City: 2008.

(166) Loss to follow-up, mortality and intermittent attendance among adults accessing HIV care in England, Wales and Northern Ireland. Proceedings of the Spring Conference of the British HIV Association: 2008.

(167) Samet JH, Freedberg KA, Stein MD, Lewis R, Savetsky J, Sullivan L et al. Trillion virion delay: time from testing positive for HIV to presentation for primary care. *Arch Intern Med* 1998; 158(7):734-740.

(168) Girardi E, Aloisi MS, Arici C, Pezzotti P, Serraino D, Balzano R et al. Delayed presentation and late testing for HIV: demographic and behavioral risk factors in a multicenter study in Italy. *J Acquir Immune Defic Syndr* 2004; 36(4):951-959.

(169) Turner BJ, Cunningham WE, Duan N, Andersen RM, Shapiro MF, Bozzette SA et al. Delayed medical care after diagnosis in a US national probability sample of persons infected with human immunodeficiency virus. *Arch Intern Med* 2000; 160(17):2614-2622.

(170) Zetola NM, Bernstein K, Ahrens K, Marcus JL, Philip S, Nieri G et al. Using surveillance data to monitor entry into care of newly diagnosed HIV-infected persons: San Francisco, 2006-2007. *BMC Public Health* 2009; 9:17.

(171) Erwin J, Morgan M, Britten N, Gray K, Peters B. Pathways to HIV testing and care by black African and white patients in London. *Sex Transm Infect* 2002; 78(1):37-39.

(172) McHenry A, Macdonald N, Sinka K, Mortimer J, Evans B. National assessment of prevalent diagnosed HIV infections. *Commun Dis Public Health* 2000; 3(4):277-281.

(173) Bellis MA, McCullagh J, Thomson R, Regan D, Syed Q, Kelly T. Inequality in funding for AIDS across England threatens regional services. *BMJ* 1997; 315(7113):950-951.

(174) Sinka K, Mortimer J, Evans B, Morgan D. Impact of the HIV epidemic in sub-Saharan Africa on the pattern of HIV in the UK. *AIDS* 2003; 17(11):1683-1690.

(175) Infuso A, Hamers F, Downs AM, Alix J. HIV reporting in western Europe : national systems and first European data. *Euro Surveill* 2000; 5(2):13-17.

(176) EuroHIV. Report on the EuroHIV 2006 survey on HIV and AIDS surveillance in the WHO European Region, Saint-Maurice: Institut de veille sanitaire, 2007. 2007. Institut de Veille Sanitaire.

(177) Bayer R, Fairchild A. The Role of Name-based Notification in Public Health and HIV Surveillance. UNAIDS/00.28E. 2000. UNAIDS.

(178) Hamers FF, Infuso A, Alix J, Downs AM. Current situation and regional perspective on HIV/AIDS surveillance in europe. *J Acquir Immune Defic Syndr* 2003; 32 Suppl 1:S39-S48.

(179) CD4 Collaborative Group. CD4 surveillance in Scotland: perspectives on severe HIV-related immunodeficiency. *AIDS* 1997; 11(12):1509-1517.

(180) Goldberg D. Surveillance of CD4 counts in known HIV-infected people--a paradigm for the era of highly active antiretroviral therapy. *Commun Dis Public Health* 2001; 4(1):4-5.

(181) Allardice GM, McMenamin JJ, Parpia T, Gibbs J, McSharry C, Whitelaw J. The recent impact of antiretroviral combination therapy on CD4 counts, AIDS and death in HIV-infected persons: routine HIV surveillance in Scotland. *International Journal of STD & AIDS* 1998; 9(10):561-566.

(182) Goldberg DJ, Emslie JA, Smyth W, Reid D. A system for surveillance of voluntary HIV testing: results of the first 2 years, 1989-1990. *AIDS* 1992; 6(5):495-500.

(183) Yirrell DL, Shaw L, Burns SM, Cameron SO, Quigg M, Campbell E et al. HIV-1 subtype in Scotland: the establishment of a national surveillance system. *Epidemiol Infect* 2004; 132(4):693-698.

(184) Li Y, Gold J, McDonald AM, Kaldor JM. Demographic pattern of AIDS in Australia, 1991 to 1993. The National HIV Surveillance Committee. *Aust N Z J Public Health* 1996; 20(4):421-425.

(185) Kaldor J, McDonald A. HIV/AIDS surveillance systems in Australia. *J Acquir Immune Defic Syndr* 2003; 32 Suppl 1:S18-S23.

(186) National HIV/AIDS Case Reporting: Surveillance Program Standard Operating Procedures. National Centre in HIV Epidemiology and Clinical Research [2007 Available from: URL:http://www.nchecr.unsw.edu.au/NCHECRweb.nsf/resources/NotiForms_3/$file/SurvSOP_Aug07.pdf

(187) Role of the National Centre in HIV Epidemiology and Clinical Research in surveillance for HIV/ AIDS, viral hepatitis and sexually transmissible infections in Australia. *Commun Dis Intell* 2004; 28(2):273-276.

(188) Semaille C. Compulsory notification of HIV infection within a new system for anonymous reporting of notifiable diseases in France. Euro Surveill 2003; 7(12).

(189) Lot F, Semaille C, Cazein F, Barin F, Pinget R, Pillonel J et al. Preliminary results from the new HIV surveillance system in France. *Euro Surveill* 2004; 9(10):34-37.

(190) Semaille C, Cazein F, Pillonel J, Lot F, Le VS, Pinget R et al. Four years of surveillance of recent HIV infections at country level, France. *Euro Surveill* 2008; 13(36).

(191) Diagnosis and reporting of HIV and AIDS in states with HIV/AIDS surveillance--United States, 1994-2000. *MMWR Morb Mortal Wkly Rep* 2002; 51(27):595-598.

(192) HIV Infection Reporting. Department of Health and Human Services: Centers for Disease Control and Prevention [2008 [cited 2009 July 17]; Available from: URL:http://www.cdc.gov/hiv/topics/surveillance/reporting.htm

(193) Glynn MK, Ling Q, Phelps R, Li J, Lee LM. Accurate monitoring of the HIV epidemic in the United States: case duplication in the national HIV/AIDS surveillance system. *J Acquir Immune Defic Syndr* 2008; 47(3):391-396.

(194) Update: public health surveillance for HIV infection--United States, 1989 and 1990. *MMWR Morb Mortal Wkly Rep* 1990; 39(47):853, 859-853, 861.

(195) Glynn MK, Lee LM, McKenna MT. The status of national HIV case surveillance, United States 2006. *Public Health Rep* 2007; 122 Suppl 1:63-71.

(196) McNaghten AD, Wolfe MI, Onorato I, Nakashima AK, Valdiserri RO, Mokotoff E et al. Improving the representativeness of behavioral and clinical surveillance for persons with HIV in the United States: the rationale for developing a population-based approach. *PLoS ONE* 2007; 2(6):e550.

(197) Smith E, Rix BA, Melbye M. Mandatory anonymous HIV surveillance in Denmark: the first results of a new system. *Am J Public Health* 1994; 84(12):1929-1932.

(198) Smith E. HIV/AIDS surveillance in Denmark: the challenges ahead. *J Acquir Immune Defic Syndr* 2003; 32 Suppl 1:S33-S38.

(199) Hamouda O. HIV/AIDS surveillance in Germany. *J Acquir Immune Defic Syndr* 2003; 32 Suppl 1:S49-S54.

(200) Brancato G, Perucci CA, Abeni DD, Sangalli M, Ippolito G, Arca M. The changing distribution of HIV infection: HIV surveillance in Lazio, Italy, 1985 through 1994. Lazio HIV Surveillance Collaborative Group. *Am J Public Health* 1997; 87(10):1654-1658.

(201) Porta D, Perucci CA, Forastiere F, De LA. Temporal trend of HIV infection: an update of the HIV surveillance system in Lazio, Italy, 1985-2000. *Eur J Public Health* 2004; 14(2):156-160.

(202) Pezzotti P, Piovesan C, Rezza G, Moro G, Serpelloni G, Ferraro P. Combining HIV and AIDS surveillance: an experience from an Italian region. *Int J Epidemiol* 1997; 26(6):1352-1358.

(203) Stichting HIV Monitoring Foundation. HIV Monitoring Foundation [2009 [cited 2009 July 14]; Available from: URL:http://www.hiv-monitoring.nl/index.jsp?ACTION=GOHOME&MID=2

(204) Polesel J, Clifford GM, Rickenbach M, Dal Maso L, Battegay M, Bouchardy C et al. Non-Hodgkin lymphoma incidence in the Swiss HIV Cohort Study before and after highly active antiretroviral therapy. *AIDS* 2008; 22(2):301-306.

(205) Engels EA, Biggar RJ, Hall HI, Cross H, Crutchfield A, Finch JL et al. Cancer risk in people infected with human immunodeficiency virus in the United States. *Int J Cancer* 2008; 123(1):187-194.

(206) Deapen D, Cockburn M, Pinder R, Lu S, Wohl AR. Population-based linkage of AIDS and cancer registries: importance of linkage algorithm. *Am J Prev Med* 2007; 33(2):134-136.

(207) Ahmed AB, Abubakar I, Delpech V, Lipman M, Boccia D, Forde J et al. The growing impact of HIV infection on the epidemiology of tuberculosis in England and Wales: 1999 2003. *Thorax* 2007; 62(8):672-676.

(208) Moore M, McCray E, Onorato IM. Cross-matching TB and AIDS registries: TB patients with HIV co-infection, United States, 1993-1994. *Public Health Rep* 1999; 114(3):269-277.

(209) Gollub EL, Trino R, Salmon M, Moore L, Dean JL, Davidson BL. Co-occurrence of AIDS and tuberculosis: results of a database "match" and investigation. *J Acquir Immune Defic Syndr Hum Retrovirol* 1997; 16(1):44-49.

(210) Libbus MK, Phillips L, Knudson J. TB-HIV Registry Matching in Missouri, 1987-1999. *Public Health Nurs* 2002; 19(6):470-474.

(211) Lewden C, Jougla E, Alioum A, Pavillon G, Lievre L, Morlat P et al. Number of deaths among HIV-infected adults in France in 2000, three-source capture-recapture estimation. *Epidemiol Infect* 2006; 134(6):1345-1352.

(212) Pezzotti P, Piovesan C, Michieletto F, Zanella F, Rezza G, Gallo G. Estimating the cumulative number of human immunodeficiency virus diagnoses by cross-linking from four different sources. *Int J Epidemiol* 2003; 32(5):778-783.

(213) Bernillon P, Lievre L, Pillonel J, Laporte A, Costagliola D. Record-linkage between two anonymous databases for a capture-recapture estimation of underreporting of AIDS cases: France 1990-1993. The Clinical Epidemiology Group from Centres d'Information et de Soins de l'Immunodeficience Humaine. *Int J Epidemiol* 2000; 29(1):168-174.

(214) Gomatam S, Carter R, Ariet M, Mitchell G. An empirical comparison of record linkage procedures. *Stat Med* 2002; 21(10):1485-1496.

(215) Herman AA, McCarthy BJ, Bakewell JM, Ward RH, Mueller BA, Maconochie NE et al. Data linkage methods used in maternally-linked birth and infant death surveillance data sets from the United States (Georgia, Missouri, Utah and Washington State), Israel, Norway, Scotland and Western Australia. *Paediatr Perinat Epidemiol* 1997; 11 Suppl 1:5-22.

(216) Blakely T, Salmond C. Probabilistic record linkage and a method to calculate the positive predictive value. *Int J Epidemiol* 2002; 31(6):1246-1252.

(217) Clark DE. Practical introduction to record linkage for injury research. *Inj Prev* 2004; 10(3):186-191.

(218) Cortina-Borja M, Cliffe S, Tookey P, Williams D, Cubitt WD, Peckham C. HIV prevalence in pregnant women in an ethnically diverse population in the UK: 1998-2002. *AIDS* 2004; 18(3):535-540.

(219) Meray N, Reitsma JB, Ravelli AC, Bonsel GJ. Probabilistic record linkage is a valid and transparent tool to combine databases without a patient identification number. *J Clin Epidemiol* 2007; 60(9):883-891.

(220) Grannis SJ, Overhage JM, McDonald CJ. Analysis of identifier performance using a deterministic linkage algorithm. *Proc AMIA Symp* 2002;305-309.

(221) Hankin C, Lyall H, Peckham C, Tookey P. Monitoring death and cancer in children born to HIV-infected women in England and Wales: use of HIV surveillance and national routine data. *AIDS* 2007; 21(7):867-869.

(222) Kendrick SW, Douglas MM, Gardner D, Hucker D. Best-link matching of Scottish health data sets. *Methods Inf Med* 1998; 37(1):64-68.

(223) McHenry A, Evans BG, Sinka K, Shaheem Z, Macdonald N, De Angelis D. Numbers of adults with diagnosed HIV infection 1996-2005--adjusted totals and extrapolations for England, Wales and Northern Ireland. *Commun Dis Public Health* 2002; 5(2):97-100.

(224) Rogers PA, Sinka KJ, Molesworth AM, Evans BG, Allardice GM. Survival after diagnosis of AIDS among adults resident in the United Kingdom in the era of multiple therapies. *Commun Dis Public Health* 2000; 3(3):188-194.

(225) Dougan S, Harris JP. Overcounting of black Africans in the UK: the problem of undetected record duplication. *Commun Dis Public Health* 2003; 6(2):147-151.

(226) Barlow D, Mears A, Gum D. Sexually Transmitted Infections: The Facts. Oxford University Press; 2006.

(227) Barton S, Roth P. Life insurance and HIV antibody testing. *BMJ* 1992; 305(6859):902-903.

(228) Byrne L. Insurers relax questions on HIV. *BMJ* 1994; 309(6951):360.

(229) Nakhaee F, McDonald A, Black D, Law M. A feasible method for linkage studies avoiding clerical review: linkage of the national HIV/AIDS surveillance databases with the National Death Index in Australia. *Aust N Z J Public Health* 2007; 31(4):308-312.

(230) Muse AG, Mikl J, Smith PF. Evaluating the quality of anonymous record linkage using deterministic procedures with the New York State AIDS registry and a hospital discharge file. *Stat Med* 1995; 14(5-7):499-509.

(231) Rice BD, McHenry A, Sinka K, Payne LJ, Baster K, Patel B et al. Prevalent diagnosed HIV in England, Wales and Northern Ireland: adjusted totals 1996 to 2001 and extrapolations to 2004. *AIDS* 2004; 18(6):927-932.

(232) The creation of a large UK-based multicentre cohort of HIV-infected individuals: The UK Collaborative HIV Cohort (UK CHIC) Study. *HIV Med* 2004; 5(2):115-124.

(233) Easterbrook PJ, Yu LM, McLean K, Hawkins D, Gazzard B. CD4 cell counts of 200 x 10(6)/1 or below in natural history studies and surveillance: is one enough or are two better? *Commun Dis Rep CDR Rev* 1997; 7(12):R179-R185.

(234) Hallett TB, Gregson S, Dube S, Garnett GP. The impact of monitoring HIV patients prior to treatment in resource-poor settings: insights from mathematical modelling. *PLoS Med* 2008; 5(3):e53.

(235) Mocroft A, Vella S, Benfield TL, Chiesi A, Miller V, Gargalianos P et al. Changing patterns of mortality across Europe in patients infected with HIV-1. EuroSIDA Study Group. *Lancet* 1998; 352(9142):1725-1730.

(236) Weatherburn P, Ssanyu-SSeruma W, Hickson FC, McLean S, Reid D. Project Nasah: An investigation into the HIV treatment information and other needs of African people with HIV resident in England. Sigma Research, NAM, National AIDS Trust, African HIV Policy Network; 2003.

(237) Hall HI, Li J, Campsmith M, Sweeny P, Lee LM. Date of first positive HIV test: reliability of information collected for HIV/AIDS surveillance in the United States. *Public Health Rep* 2005; 120(1):89-95.

(238) Postma MJ, Beck EJ, Hankins CA, Mandalia S, Jager JC, de Jong-van den Berg LT et al. Cost effectiveness of expanded antenatal HIV testing in London. *AIDS* 2000; 14(15):2383-2389.

(239) Chadborn TR, Baster K, Delpech VC, Sabin CA, Sinka K, Rice BD et al. No time to wait: how many HIV-infected homosexual men are diagnosed late and consequently die?(England and Wales, 1993-2002). *AIDS* 2005; 19(5):513-520.

(240) Chadborn TR, Delpech VC, Sabin CA, Sinka K, Evans BG. The late diagnosis and consequent short-term mortality of HIV-infected heterosexuals (England and Wales, 2000-2004). *AIDS* 2006; 20(18):2371-2379.

(241) Department of Health. The national strategy for sexual health and HIV: Implementation Action Plan. Crown Copyright; 2002.

(242) Hoover DR, Graham NM, Chen B, Taylor JM, Phair J, Zhou SY et al. Effect of CD4+ cell count measurement variability on staging HIV-1

infection. *Journal of Acquired Immune Deficiency Syndromes* 1992; 5(8):794-802.

(243) Raboud JM, Haley L, Montaner JS, Murphy C, Januszewska M, Schechter MT. Quantification of the variation due to laboratory and physiologic sources in CD4 lymphocyte counts of clinically stable HIV-infected individuals. *Journal of Acquired Immune Deficiency Syndromes & Human Retrovirology* 1995; 10:Suppl-73.

(244) Paglieroni TG, Holland PV. Circannual variation in lymphocyte subsets, revisited. *Transfusion (Paris)* 1994; 34(6):512-516.

(245) von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol* 2008; 61(4):344-349.

(246) Sabin CA, Phillips AN. Cohort studies in sexual health. *Sex Transm Infect* 2001; 77(3):174-178.

(247) Porter K, Johnson AM, Phillips AN, Darbyshire JH. The practical significance of potential biases in estimates of the AIDS incubation period distribution in the UK register of HIV seroconverters. *AIDS* 1999; 13(14):1943-1951.

(248) Griffin JT, Fraser C, Gras L, de Wolf F, Ghani AC. The effect on treatment comparisons of different measurement frequencies in human immunodeficiency virus observational databases. *Am J Epidemiol* 2006; 163(7):676-683.

(249) Lau B, Gange SJ, Moore RD. Interval and clinical cohort studies: epidemiological issues. *AIDS Res Hum Retroviruses* 2007; 23(6):769-776.

(250) Lanoy E, Lewden C, Lievre L, Tattevin P, Boileau J, Aouba A et al. How does loss to follow-up influence cohort findings on HIV infection? A joint analysis of the French hospital database on HIV, Mortalite 2000 survey and death certificates. *HIV Med* 2009.

(251) Touloumi G, Babiker AG, Pocock SJ, Darbyshire JH. Impact of missing data due to drop-outs on estimators for rates of change in longitudinal studies: a simulation study. *Stat Med* 2001; 20(24):3715-3728.

(252) Touloumi G, Pocock SJ, Babiker AG, Darbyshire JH. Impact of missing data due to selective dropouts in cohort studies and clinical trials. *Epidemiology* 2002; 13(3):347-355.

(253) Roy J, Lin X. Missing covariates in longitudinal data with informative dropouts: bias analysis and inference. *Biometrics* 2005; 61(3):837-846.

(254) Raboud JM, Montaner JS, Thorne A, Singer J, Schechter MT. Impact of missing data due to dropouts on estimates of the treatment effect in a randomized trial of antiretroviral therapy for HIV-infected individuals. Canadian HIV Trials Network A002 Study Group. *J Acquir Immune Defic Syndr Hum Retrovirol* 1996; 12(1):46-55.

(255) Shepherd BE, Sterling TR, Moore RD, Raffanti SP, Hulgan T. Cross-cohort heterogeneity encountered while validating a model for HIV disease progression among antiretroviral initiators. *J Clin Epidemiol* 2008.

(256) Ioannidis JP, Bassett R, Hughes MD, Volberding PA, Sacks HS, Lau J. Predictors and impact of patients lost to follow-up in a long-term randomized trial of immediate versus deferred antiretroviral treatment. *J Acquir Immune Defic Syndr Hum Retrovirol* 1997; 16(1):22-30.

(257) Anglaret X, Toure S, Gourvellec G, Tchehy A, Zio L, Zaho M et al. Impact of vital status investigation procedures on estimates of survival in cohorts of HIV-infected patients from Sub-Saharan Africa. *J Acquir Immune Defic Syndr* 2004; 35(3):320-323.

(258) Lomax N, Curtis H, Johnson M. A national review of assessment and monitoring of HIV-infected patients. *HIV Med* 2009; 10(2):125-128.

(259) Gazzard BG. British HIV Association Guidelines for the treatment of HIV-1-infected adults with antiretroviral therapy 2008. *HIV Med* 2008; 9(8):563-608.

(260) Pacheco AG, Saraceni V, Tuboi SH, Moulton LH, Chaisson RE, Cavalcante SC et al. Validation of a hierarchical deterministic record-linkage algorithm using data from 2 different cohorts of human immunodeficiency virus-infected persons and mortality databases in Brazil. *Am J Epidemiol* 2008; 168(11):1326-1332.

(261) Brenner H, Schmidtmann I, Stegmaier C. Effects of record linkage errors on registry-based follow-up studies. *Stat Med* 1997; 16(23):2633-2643.

(262) Nedelman J, Burns A, Cleary J, Gordon D, Vernon P, Lawrence CE. Modelling length bias in a longitudinally-linked record system of HIV cases. *Stat Med* 1991; 10(3):423-431.

(263) Hunt JR, White E. Retaining and tracking cohort study members. *Epidemiol Rev* 1998; 20(1):57-70.

(264) Health Protection Agency, Health Protection Scotland, National Public Health Services for Wales, CDSC Northern Ireland & CRDHB. Shooting Up: Infections among injecting drug users in the United Kingdom 2006: an update. 2007. London, Health Protection Agency. 2006.

(265) Gerver SM, Chadborn TR, Ibrahim F, Vatsa B, Delpech VC, Easterbrook PJ. High rate of loss to clinical follow up among African HIV-infected patients attending a London clinic: a retrospective analysis of a clinical cohort. *J Int AIDS Soc* 2010; 13:29.

(266) Creighton S, Sethi G, Edwards SG, Miller R. Dispersal of HIV positive asylum seekers: national survey of UK healthcare providers. *BMJ* 2004; 329(7461):322-323.

(267) Thomas F, Aggleton P, Anderson J. "If I cannot access services, then there is no reason for me to test": the impacts of health service charges on HIV testing and treatment amongst migrants in England. *AIDS Care* 2010; 22(4):526-531.

(268) National AIDS Trust. Speaking Out, Changing Lives: HIV and the UK Asylum Pathway. 1-18. 8. London, National AIDS Trust.

(269) Department of Health and Refugee Council. Caring for Dispersed Asylum Seekers: A Resource Pack. Crown; 2003.

(270) Mocroft A, Kirk O, Aldins P, Chies A, Blaxhult A, Chentsova N et al. Loss to follow-up in an international, multicentre observational study. *HIV Med* 2008; 9(5):261-269.

(271) Ulett KB, Willig JH, Lin HY, Routman JS, Abroms S, Allison J et al. The therapeutic implications of timely linkage and early retention in HIV care. *AIDS Patient Care STDS* 2009; 23(1):41-49.

(272) Andersen JW, Fass R, van der HC. Factors associated with early study discontinuation in AACTG studies, DACS 200. *Contemp Clin Trials* 2007; 28(5):583-592.

(273) Lanoy E, Mary-Krause M, Tattevin P, Dray-Spira R, Duvivier C, Fischer P et al. Predictors identified for losses to follow-up among HIV-seropositive patients. *J Clin Epidemiol* 2006; 59(8):829-835.

(274) Robison LS, Westfall AO, Mugavero MJ, Kempf MC, Cole SR, Allison JJ et al. Short-term discontinuation of HAART regimens more common in vulnerable patient populations. *AIDS Res Hum Retroviruses* 2008; 24(11):1347-1355.

(275) Nacher M, El Guedj M, Vaz T, Nasser V, Randrianjohany A, Alvarez F et al. Risk factors for follow-up interruption of HIV patients in French Guiana. *Am J Trop Med Hyg* 2006; 74(5):915-917.

(276) Lebouche B, Yazdanpanah Y, Gerard Y, Sissoko D, Ajana F, Alcaraz I et al. Incidence rate and risk factors for loss to follow-up in a French clinical cohort of HIV-infected patients from January 1985 to January 1998. *HIV Med* 2006; 7(3):140-145.

(277) Dudley J, Jin S, Hoover D, Metz S, Thackeray R, Chmiel J. The Multicenter AIDS Cohort Study: retention after 9 1/2 years. *Am J Epidemiol* 1995; 142(3):323-330.

(278) Johnson MO, Dilworth SE, Neilands TB, Chesney MA, Rotheram-Borus MJ, Remien RH et al. Predictors of attrition among high risk HIV-infected participants enrolled in a multi-site prevention trial. *AIDS Behav* 2008; 12(6):974-977.

(279) Israelski D, Gore-Felton C, Power R, Wood MJ, Koopman C. Sociodemographic characteristics associated with medical appointment adherence among HIV-seropositive patients seeking treatment in a county outpatient facility. *Prev Med* 2001; 33(5):470-475.

(280) Hessol NA, Schneider M, Greenblatt RM, Bacon M, Barranday Y, Holman S 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):563-573.

(281) Kissinger P, Cohen D, Brandon W, Rice J, Morse A, Clark R. Compliance with public sector HIV medical care. *J Natl Med Assoc* 1995; 87(1):19-24.

(282) Health Protection Agency. Testing Times. HIV and other Sexually Transmitted Infections in the United Kingdom: 2007. 2007. London, Health Protection Agency Centre for Infections.

(283) Pearshouse R. U.N. Secretary-General calls for an end to criminal law approaches to vulnerable populations. *HIV AIDS Policy Law Rev* 2008; 13(1):45.

(284) Clarke S, Delamere S, McCullough L, Hopkins S, Bergin C, Mulcahy F. Assessing limiting factors to the acceptance of antiretroviral therapy in a large cohort of injecting drug users. *HIV Med* 2003; 4(1):33-37.

(285) Health Protection Agency, National Public Health Service for Wales, SCIEH, ISD, CDSC Northern Ireland and the UASSG. Mapping the Issues. HIV and other Sexually Transmitted Infections in the United Kingdom: 2005. London: Health Protection Agency; 2005.

(286) Solomon L, Stein M, Flynn C, Schuman P, Schoenbaum E, Moore J et al. Health services use by urban women with or at risk for HIV-1 infection: the HIV Epidemiology Research Study (HERS). *J Acquir Immune Defic Syndr Hum Retrovirol* 1998; 17(3):253-261.

(287) Hope SC, Williams AE, Barton SE, Asboe D. What do patients attending HIV and GUM outpatient clinics want from service providers? Results from a large-scale consultation exercise in west London. *Int J STD AIDS* 2001; 12(11):733-738.

(288) Brown DM, Thorne JE, Foster GL, Duncan JL, Brune LM, Munana A et al. Factors affecting attrition in a longitudinal study of patients with AIDS. *AIDS Care* 2006; 18(7):821-829.

(289) Bakken S, Holzemer WL, Brown MA, Powell-Cope GM, Turner JG, Inouye J et al. Relationships between perception of engagement with health care provider and demographic characteristics, health status, and adherence to therapeutic regimen in persons with HIV/AIDS. *AIDS Patient Care STDS* 2000; 14(4):189-197.

(290) Braitstein P, Brinkhof MW, Dabis F, Schechter M, Boulle A, Miotti P et al. Mortality of HIV-1-infected patients in the first year of antiretroviral therapy: comparison between low-income and high-income countries. *Lancet* 2006; 367(9513):817-824.

(291) Catz SL, McClure JB, Jones GN, Brantley PJ. Predictors of outpatient medical appointment attendance among persons with HIV. *AIDS Care* 1999; 11(3):361-373.

(292) Arici C, Ripamonti D, Maggiolo F, Rizzi M, Finazzi MG, Pezzotti P et al. Factors associated with the failure of HIV-positive persons to return for scheduled medical visits. *HIV Clin Trials* 2002; 3(1):52-57.

(293) Lemly D, Mandelbrot L, Meier F, Firtion G, Matheron S, Jeantils V et al. Factors related to medical appointment attendance after childbirth among HIV-infected women in the Paris region. *AIDS Care* 2007; 19(3):346-354.

(294) Chinouya M, Davidson O. The Padare Project: Assessing health-related knowledge, attitudes and behaviours of HIV-positive Africans accessing services in north central London. London: African HIV Policy Network; 2003.

(295) Dodds C, Hickson F, Chinouya M, Chwaula J, Weatherburn P. The Knowledge, the Will and the Power: A plan of action to meet the HIV prevention needs of Africans living in England. Portsmouth: Sigma Research; 2008.

(296) Dodds C, Hickson F, Weatherburn P, Reid D, Hammond G, Jessup K et al. Bass Line 2007 survey: Assessing the sexual HIV prevention needs of African people in England. Portsmouth: Sigma Research; 2008.

(297) Chinouya M, Ssanyu-SSeruma W, Kwok A. The SHIBAH Report: A study of sexual health issues affecting black Africans living with HIV in Lambeth, Southwark and Lewisham. London: Health First: Specialists in Health Promotion for Lambeth, Southwark and Lewisham; 2003.

(298) Coleman S, Boehmer U, Kanaya F, Grasso C, Tan J, Bradford J. Retention challenges for a community-based HIV primary care clinic and

implications for intervention. *AIDS Patient Care STDS* 2007; 21(9):691-701.

(299) Phillips A. Short-term risk of AIDS according to current CD4 cell count and viral load in antiretroviral drug-naive individuals and those treated in the monotherapy era. *AIDS* 2004; 18(1):51-58.

(300) Easterbrook PJ, Yu LM, Goetghebeur E, Boag F, McLean K, Gazzard B. Ten-year trends in CD4 cell counts at HIV and AIDS diagnosis in a London HIV clinic. *AIDS* 2000; 14(5):561-571.

(301) Ndiaye B, Salleron J, Vincent A, Bataille P, Bonnevie F, Choisy P et al. Factors associated with presentation to care with advanced HIV disease in Brussels and Northern France: 1997-2007. *BMC Infect Dis* 2011; 11:11.

(302) Milberg J, Sharma R, Scott F, Conviser R, Marconi K, Parham D. Factors associated with delays in accessing HIV primary care in rural Arkansas. *AIDS Patient Care STDS* 2001; 15(10):527-532.

(303) Bamford LP, Ehrenkranz PD, Eberhart MG, Shpaner M, Brady KA. Factors associated with delayed entry into primary HIV medical care after HIV diagnosis. *AIDS* 2010; 24(6):928-930.

(304) Plitt SS, Mihalicz D, Singh AE, Jayaraman G, Houston S, Lee BE. Time to Testing and Accessing Care among a Population of Newly Diagnosed Patients with HIV with a High Proportion of Canadian Aboriginals, 1998-2003. *AIDS Patient Care STDS* 2009.

(305) Torian LV, Wiewel EW, Liu KL, Sackoff JE, Frieden TR. Risk factors for delayed initiation of medical care after diagnosis of human immunodeficiency virus. *Arch Intern Med* 2008; 168(11):1181-1187.

(306) Girardi E, Aloisi MS, Arici C, Pezzotti P, Serraino D, Balzano R et al. Delayed presentation and late testing for HIV: demographic and behavioral risk factors in a multicenter study in Italy. *J Acquir Immune Defic Syndr* 2004; 36(4):951-959.

(307) Page LC, Goldbaum G, Kent JB, Buskin SE. Access to regular HIV care and disease progression among black African immigrants. *J Natl Med Assoc* 2009; 101(12):1230-1236.

(308) Krawczyk CS, Funkhouser E, Kilby JM, Kaslow RA, Bey AK, Vermund SH. Factors associated with delayed initiation of HIV medical care among infected persons attending a southern HIV/AIDS clinic. *South Med J* 2006; 99(5):472-481.

(309) Stein MD, Crystal S, Cunningham WE, Ananthanarayanan A, Andersen RM, Turner BJ et al. Delays in seeking HIV care due to competing caregiver responsibilities. *Am J Public Health* 2000; 90(7):1138-1140.

(310) Randall M, Barroso J. Delayed pursuit of health care among HIV-positive gay men enrolled in a longitudinal research study. *J Assoc Nurses AIDS Care* 2002; 13(4):23-31.

(311) Raveis VH, Siegel K, Gorey E. Factors associated with HIV-infected women's delay in seeking medical care. *AIDS Care* 1998; 10(5):549-562.

(312) Kumar A, Kilaru KR, Kumari G, Forde S, Waterman I. Follow-up of HIV-infected women diagnosed by antenatal screening in Barbados from 1996-2004. *AIDS Patient Care STDS* 2008; 22(9):715-721.

(313) Ickovics JR, Forsyth B, Ethier KA, Harris P, Rodin J. Delayed entry into health care for women with HIV disease. *AIDS Patient Care STDS* 1996; 10(1):21-24.

(314) Reed JB, Hanson D, McNaghten AD, Bertolli J, Teshale E, Gardner L et al. HIV testing factors associated with delayed entry into HIV medical care among HIV-infected persons from eighteen states, United States, 2000-2004. *AIDS Patient Care STDS* 2009; 23(9):765-773.

(315) Reekie J, Mocroft A, Sambatakou H, Machala L, Chiesi A, van Lunzen J et al. Does less frequent routine monitoring of patients on a stable, fully suppressed cART regimen lead to an increased risk of treatment failure? *AIDS* 2008; 22(17):2381-2390.

(316) Harte D, Dosekun O, Sethi G, Chadborn T, de RA, Copas A et al. Immunosuppression among HIV-1-positive patients attending for care: experience from two large HIV centres in the United Kingdom. *HIV Med* 2010; 11(2):114-120.

(317) Turner BJ, Hecht FM, Ismail RB. CD4+ T-lymphocyte measures in the treatment of individuals infected with human immunodeficiency virus type 1. A review for clinical practitioners. *Arch Intern Med* 1994; 154(14):1561-1573.

(318) Kimmel AD, Goldie SJ, Walensky RP, Losina E, Weinstein MC, Paltiel AD et al. Optimal frequency of CD4 cell count and HIV RNA monitoring prior to initiation of antiretroviral therapy in HIV-infected patients. *Antivir Ther* 2005; 10(1):41-52.

(319) Morris S, Gray A, Noone A, Wiseman M, Jathanna S. The costs and effectiveness of surveillance of communicable disease: a case study of HIV and AIDS in England and Wales. *J Public Health Med* 1996; 18(4):415-422.

(320) Gupta SB, Gilbert RL, Brady AR, Livingstone SJ, Evans BG. CD4 cell counts in adults with newly diagnosed HIV infection: results of surveillance in England and Wales, 1990-1998. CD4 Surveillance Scheme Advisory Group. *AIDS* 2000; 14(7):853-861.

(321) Saul J, Erwin J, Bruce JC, Peters B. Ethnic and demographic variations in HIV/AIDS presentation at two London referral centres 1995-9. *Sex Transm Infect* 2000; 76(3):215.

(322) Del Amo J, Petruckevitch A, Phillips AN, Johnson AM, Stephenson JM, Desmond N et al. Spectrum of disease in Africans with AIDS in London. *AIDS* 1996; 10(13):1563-1569.

(323) Sethi G, Lacey CJ, Fenton KA, Williams IG, Fox E, Sabin CA et al. South Asians with HIV in London: is it time to rethink sexual health service delivery to meet the needs of heterosexual ethnic minorities? *Sex Transm Infect* 2004; 80(1):75-76.

(324) Barry SM, Lloyd-Owen SJ, Madge SJ, Cozzi-Lepri A, Evans AJ, Phillips AN et al. The changing demographics of new HIV diagnoses at a London centre from 1994 to 2000. *HIV Med* 2002; 3(2):129-134.

(325) Curtis H, Sabin CA, Johnson MA, British HIV Association Clinical Audit Committee. Findings from the first national clinical audit of treatment for people with HIV. *HIV Med* 2003; 4(1):11-17.

(326) Burns FM, Fakoya AO, Copas AJ, French PD. Africans in London continue to present with advanced HIV disease in the era of highly active antiretroviral therapy. *AIDS* 2001; 15(18):2453-2455.

(327) Unlinked Anonymous Surveys Steering Group. Prevalence of HIV and hepatitis infections in the United Kingdom 2001. London: Department of Health; 2002.

(328) Weinhardt LS, Carey MP, Johnson BT, Bickham MS. Effects of HIV counseling and Testing on Sexual Risk Behaviour: A meta-Analytic Review of Published Research, 1985-1997. *Am J Public Health* 1999; 89:1397-1405.

(329) Leaity S, Sherr L, Wells H, Evans A, Miller R, Johnson M et al. Repeat HIV testing: high-risk behaviour or risk reduction strategy? *AIDS* 2000; 14(5):547-552.

(330) Sterling TR, Chaisson RE, Keruly J, Moore RD. Improved outcomes with earlier initiation of highly active antiretroviral therapy among human immunodeficiency virus-infected patients who achieve durable virologic suppression: longer follow-up of an observational cohort study. *J Infect Dis* 2003; 188(11):1659-1665.

(331) Krentz HB, Auld MC, Gill MJ. The high cost of medical care for patients who present late (CD4 <200 cells/microL) with HIV infection. *HIV Med* 2004; 5(2):93-98.

(332) Marks G, Crepaz N, Janssen RS. Estimating sexual transmission of HIV from persons aware and unaware that they are infected with the virus in the USA. *AIDS* 2006; 20(10):1447-1450.

(333) Donaldson L, Department of Health. On the state of the public health: Annual report of the Chief Medical Officer 2003. 32189 ed. Crown; 2004.

(334) Girardi E, Arici C, Ferrara M, Ripamonti D, Aloisi MS, Alessandrini A et al. Estimating duration of HIV infection with CD4 cell count and HIV-1 RNA at presentation. *AIDS* 2001; 15(16):2201-2203.

(335) Girardi E, Sabin CA, Monforte AD. Late diagnosis of HIV infection: epidemiological features, consequences and strategies to encourage earlier testing. *J Acquir Immune Defic Syndr* 2007; 46 Suppl 1:S3-S8.

(336) Gillieatt SJ, Mallal SA, French MA, Dawkins RL. Epidemiology of late presentation of HIV infection in Western Australia. *Med J Aust* 1992; 157(2):117-118.

(337) Antinori A, Coenen T, Costagiola D, Dedes N, Ellefson M, Gatell J et al. Late presentation of HIV infection: a consensus definition. *HIV Med* 2011; 12(1):61-64.

(338) Sabin CA, Schwenk A, Johnson MA, Gazzard B, Fisher M, Walsh J et al. Late diagnosis in the HAART era: proposed common definitions and associations with mortality. *AIDS* 2010; 24(5):723-727.

(339) Adler A, Mounier-Jack S, Coker RJ. Late diagnosis of HIV in Europe: definitional and public health challenges. *AIDS Care* 2008;1-10.

(340) Hu DJ, Byers R, Jr., Fleming PL, Ward JW. Characteristics of persons with late AIDS diagnosis in the United States. *Am J Prev Med* 1995; 11(2):114-119.

(341) Reedijk M, Bindels PJ, Wigersma L. Late presentation of HIV infected patients at Amsterdam outpatient HIV clinics. *Sex Transm Infect* 1999; 75(6):448.

(342) Mounier-Jack S, Nielsen S, Coker RJ. HIV testing strategies across European countries. *HIV Med* 2008; 9 Suppl 2:13-19.

(343) Sabharwal CJ, Sepkowitz K, Mehta R, Shepard C, Bodach S, Torian L et al. Impact of Accelerated Progression to AIDS on Public Health Monitoring of Late HIV Diagnosis. *AIDS Patient Care STDS* 2011.

(344) Poznansky MC, Coker R, Skinner C, Hill A, Bailey S, Whitaker L et al. HIV positive patients first presenting with an AIDS defining illness: characteristics and survival. *BMJ* 1995; 311(6998):156-158.

(345) Couturier E, Schwoebel V, Michon C, Hubert JB, Delmas MC, Morlat P et al. Determinants of delayed diagnosis of HIV infection in France, 1993-1995. *AIDS* 1998; 12(7):795-800.

(346) Girardi E, Sampaolesi A, Gentile M, Nurra G, Ippolito G. Increasing proportion of late diagnosis of HIV infection among patients with AIDS in Italy following introduction of combination antiretroviral therapy. *Journal of Acquired Immune Deficiency Syndromes: JAIDS* 2000; 25(1):71-76.

(347) Hocking JS, Rodger AJ, Rhodes DG, Crofts N. Late presentation of HIV infection associated with prolonged survival following AIDS diagnosis--characteristics of individuals. *International Journal of STD & AIDS* 2000; 11(8):503-508.

(348) Castilla J, Sobrino P, De La FL, Noguer I, Guerra L, Parras F. Late diagnosis of HIV infection in the era of highly active antiretroviral therapy: consequences for AIDS incidence. *AIDS* 2002; 16(14):1945-1951.

(349) Schwarcz S, Hsu L, Dilley JW, Loeb L, Nelson K, Boyd S. Late Diagnosis of HIV Infection: Trends, Prevalence, and Characteristics of Persons Whose HIV Diagnosis Occurred Within 12 Months of Developing AIDS. *J Acquir Immune Defic Syndr* 2006; 43(4):491-494.

(350) Castelnuovo B, Chiesa E, Rusconi S, Adorni F, Bongiovanni M, Melzi S et al. Declining incidence of AIDS and increasing prevalence of AIDS presenters among AIDS patients in Italy. *European Journal of Clinical Microbiology & Infectious Diseases* 2003; 22(11):663-669.

(351) Korner H. Late HIV diagnosis of people from culturally and linguistically diverse backgrounds in Sydney: the role of culture and community. *AIDS Care* 2007; 19(2):168-178.

(352) Lazarus JV, Jurgens R, Weait M, Phillips A, Hows J, Gatell J et al. Overcoming obstacles to late presentation for HIV infection in Europe. *HIV Med* 2011; 12(4):246-249.

(353) Gilbart VL, Dougan S, Sinka K, Evans BG. Late diagnosis of HIV infection among individuals with low, unrecognised or unacknowledged risks in England, Wales and Northern Ireland. *AIDS Care* 2006; 18(2):133-139.

(354) Thanawuth N, Chongsuvivatwong V. Late HIV diagnosis and delay in CD4 count measurement among HIV-infected patients in Southern Thailand. *AIDS Care* 2008; 20(1):43-50.

(355) Crowe SM, Carlin JB, Stewart KI, Lucas R, Hoy JF. Predictive value of CD4 lymphocyte numbers for the development of opportunistic infections and malignancies in HIV-infected persons. *J of Acq Imm Defic Synd* 1991; 4(8):770-776.

(356) Hanson DL, Chu SY, Farizo KM, Ward JW. Distribution of CD4+ T lymphocytes at diagnosis of acquired immunodeficiency syndrome-defining and other human immunodeficiency virus-related illnesses. The Adult and Adolescent Spectrum of HIV Disease Project Group. *Arch Intern Med* 1995; 155(14):1537-1542.

(357) Moore RD, Chaisson RE. Natural history of opportunistic disease in an HIV-infected urban clinical cohort. *Ann Intern Med* 1996; 124(7):633-642.

(358) Mocroft A, Johnson MA, Phillips AN. Factors affecting survival in patients with the acquired immunodeficiency syndrome. *AIDS* 1996; 10(10):1057-1065.

(359) Grabar S, Lanoy E, Allavena C, Mary-Krause M, Bentata M, Fischer P 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):246-256.

(360) Achhra AC, Amin J, Law MG, Emery S, Gerstoft J, Gordin FM et al. Immunodeficiency and the risk of serious clinical endpoints in a well studied cohort of treated HIV-infected patients. *AIDS* 2010; 24(12):1877-1886.

(361) van Sighem AI, van de Wiel MA, Ghani AC, Jambroes M, Reiss P, Gyssens IC et al. Mortality and progression to AIDS after starting highly active antiretroviral therapy. *AIDS* 2003; 17(15):2227-2236.

(362) European guidelines for the clinical management and treatment of HIV-infected adults in Europe. *AIDS* 2003; 17 Suppl 2:S3-26.

(363) Samet JH, Freedberg KA, Savetsky JB, Sullivan LM, Stein MD. Understanding delay to medical care for HIV infection: the long-term non-presenter. *AIDS* 2001; 15(1):77-85.

(364) Smith CJ, Sabin CA, Lampe FC, Kinloch-de-Loes S, Gumley H, Carroll A et al. The potential for CD4 cell increases in HIV-positive individuals who control viraemia with highly active antiretroviral therapy. *AIDS* 2003; 17(7):963-969.

(365) Mocroft A, Ledergerber B, Katlama C, Kirk O, Reiss P, d'Arminio MA et al. Decline in the AIDS and death rates in the EuroSIDA study: an observational study. *Lancet* 2003; 362(9377):22-29.

(366) PHLS D&PatSID5CG. Sexually Transmitted Infections in the UK: New Episodes seen at Genitourinary Medicine Clinics, 1991 to 2001. London: Public Health Laboratory Service; 2002.

(367) Hogg RS, Yip B, Chan KJ, Wood E, Craib KJ, O'Shaughnessy MV et al. Rates of disease progression by baseline CD4 cell count and viral load after initiating triple-drug therapy. *JAMA* 2001; 286(20):2568-2577.

(368) Sterling TR, Chaisson RE, Moore RD. Initiation of highly active antiretroviral therapy at CD4+ T lymphocyte counts of >350 cells/mm3: disease progression, treatment durability, and drug toxicity. *Clin Infect Dis* 2003; 36(6):812-815.

(369) PHLS Communicable Disease Surveillance Centre, ICH (London), SCIEH. HIV & AIDS in the United Kingdom 2001: An update November 2002. London: Public Health Laboratory Service; 2002.

(370) CASCADE (Concerted Action on SeroConversion to AIDS and Death in Europe) Collaboration. Is the time from HIV seroconversion a determinant of the risk of AIDS after adjustment for updated CD4 cell counts? *Journal of Acquired Immune Deficiency Syndromes: JAIDS* 2001; 28(2):158-165.

(371) Sullivan AK, Curtis H, Sabin CA, Johnson MA, on behalf of BHIVA and the BHIVA audit subcommittee. National Review of Newly Diagnosed , Peter Keogh, Peter Weatherburn, David Reid. Managin HIV Infections. (submitted for publication). *BMJ* 2004.

(372) Manavi K, McMillan A, Ogilvie M, Scott G. Heterosexual men and women with HIV test positive at a later stage of infection than homo- or bisexual men. *Int J STD AIDS* 2004; 15(12):811-814.

(373) Laurie Hendersong uncertainty: Risk and unprotected anal intercourse among gay men who do not know their HIV status. ISBN: 1 872956 63 7. 2001. London, Sigma Research.

(374) Chadborn TR, McGarrigle CA, Waight PA, Fenton KA. Trends in, and determinants of, HIV testing at genitourinary medicine clinics and general practice in England, 1990-2000. *Sex Transm Infect* 2004; 80(2):145-150.

(375) Public Health Laboratory Service D&PatSID5CG. Sexually Transmitted Infections in the UK: New Episodes seen at Genitourinary Medicine Clinics, 1991 to 2001. London: Public Health Laboratory Service; 2002.

(376) Murphy G, Charlett A, Brown A, Gill ON, Parry JV. Is HIV incidence increasing in homo/bisexual men attending GUM clinics in England, Wales and Northern Ireland? *Commun Dis Public Health* 2004; **7**(1):11-14.

(377) Report of a Working Group (Chairman: Professor N E Day) convened by the Director of the Public Health Laboratory Service on behalf of the Chief Medical Officers. The incidence and prevalence of AIDS and prevalence of other severe HIV disease in England and Wales for 1995

to 1999: projections using data to the end of 1994. *CDR Review* 1996; 6(1):R1-R24.

(378) Health Protection Agency. Sexually Transmitted Infections and Men who have Sex with Men in the UK: 2008 Report. 2008. London, Health Protection Agency Centre for Infections.

(379) Health Protection Agency. Sexually Transmitted Infections in black African and black Caribbean communities in the UK: 2008 report. 2008. London, Health Protection Agency Centre for Infections.

(380) Fenton KA, Chinouya M, Davidson O, Copas A. HIV testing and high risk sexual behaviour among London's migrant African communities: a participatory research study. *Sex Transm Infect* 2002; 78(4):241-245.

(381) McCoy SI, Miller WC, MacDonald PD, Hurt CB, Leone PA, Eron JJ et al. Barriers and facilitators to HIV testing and linkage to primary care: narratives of people with advanced HIV in the Southeast. *AIDS Care* 2009; 21(10):1313-1320.

(382) Burns FM, Imrie JY, Nazroo J, Johnson AM, Fenton KA. Why the(y) wait? Key informant understandings of factors contributing to late presentation and poor utilization of HIV health and social care services by African migrants in Britain. *AIDS Care* 2007; 19(1):102-108.

(383) Elford J, Ibrahim F, Bukutu C, Anderson J. HIV-Related Discrimination Reported by People Living with HIV in London, UK. *AIDS Behav* 2008; 12(2):255-264.

(384) Dodds C, Keogh P. Criminal prosecutions for HIV transmission: people living with HIV respond. *Int J STD AIDS* 2006; 17(5):315-318.

(385) Kellock DJ, Rogstad KE. Attitudes to HIV testing in general practice. *Int J STD AIDS* 1998; 9(5):263-267.

(386) Department of Health. Proposed Amendments to the NHS (Charges to Overseas Visitors) Regulations 1989: A Consultation. London: Crown; 2003.

(387) Select Committee on Health. Third Report. 2005. London, The Stationery Office of the United Kingdom Parliament.

(388) Rogstad KE. HIV testing for patients attending general medical services: concise guidelines. *Clin Med* 2004; 4(2):136-139.

(389) Peck L, Ferenczi E, Burns F, Cosgrove C, Brown M. Barriers to targeted HIV testing on an acute admissions unit: evaluation of the UK guideline. *QJM* 2010; 103(3):147-151.

(390) Yazdanpanah Y, Sloan CE, Charlois-Ou C, Le VS, Semaille C, Costagliola D et al. Routine HIV screening in France: clinical impact and cost-effectiveness. *PLoS ONE* 2010; 5(10):e13132.

(391) Delpierre C, Dray-Spira R, Cuzin L, Marchou B, Massip P, Lang T et al. Correlates of late HIV diagnosis: implications for testing policy. *Int J STD AIDS* 2007; 18(5):312-317.

(392) Gandhi NR, Skanderson M, Gordon KS, Concato J, Justice AC. Delayed presentation for human immunodeficiency virus (HIV) care among veterans: a problem of access or screening? *Med Care* 2007; 45(11):1105-1109.

(393) Sanders GD, Bayoumi AM, Sundaram V, Bilir-S-Pinar, Neukermans CP, Rydzak CE et al. Cost-effectiveness of screening for HIV in the era of highly active antiretroviral therapy. *N Engl J Med* 2005; 352(6):570-585.

(394) Health Protection Agency. HIV in the United Kingdom: 2008 Report. 2008. London, Health Protection Agency Centre for Infections.

(395) Munro HL, Lowndes CM, Daniels DG, Sullivan AK, Robinson AJ. National study of HIV testing in men who have sex with men attending genitourinary clinics in the United Kingdom. *Sex Transm Infect* 2008; 84(4):265-270.

(396) New guidelines for HIV testing and areas where wider HIV testing should be considered. Health Protection Agency Centre for Infections [2009 [cited 9 A.D. June 28]; Available from: URL:http://www.hpa.org.uk/web/HPAweb&HPAwebStandard/HPAweb_C/1201767906579

(397) Department of Health, RefugeeCouncil. Caring for Dispersed Asylum Seekers: A Resource Pack. Crown; 2003.

(398) Home Office. Integration Matters: A National Strategy for Refugee Integration. London: IND Corporate Communications; 2005.

(399) Home Office. Control of Immigration: Statistics, UK, 2002. Crown; 2005.

(400) Home Office. Home Office Statistical Bulletin: Asylum Statistics, United Kingdom 2007. Home Office [2008 [cited 9 A.D. June 28]; Available from: URL:http://www.homeoffice.gov.uk/rds/pdfs08/hosb1108.pdf

(401) Rajamanoharan S, Monteiro EF, Maw R, Carne CA, Robinson A. Genitourinary medicine/HIV services for persons with insecure immigration or seeking asylum in the United Kingdom: a British Co-operative Clinical Group survey. *Int J STD AIDS* 2004; 15(8):509-514.

(402) Cook PA, Downing J, Jones AM, Rimmer P, Ashton M, Syed Q et al. HIV and AIDS in the north west of England: mid year 2004. North West HIV/AIDS Monitoring Unit; 2004.

(403) Wolbers M, Bucher HC, Furrer H, Rickenbach M, Cavassini M, Weber R et al. Delayed diagnosis of HIV infection and late initiation of antiretroviral therapy in the Swiss HIV Cohort Study. *HIV Med* 2008; 9(6):397-405.

(404) Santos J, Palacios R, Gutierrez M, Grana M, de la Torre J, Salgado F et al. HIV infection in the era of highly active antiretroviral therapy: the Malaga Study. *Int J STD AIDS* 2004; 15(9):594-596.

(405) Lee JH, Kim GJ, Choi BS, Hong KJ, Heo MK, Kim SS et al. Increasing late diagnosis in HIV infection in South Korea: 2000-2007. *BMC Public Health* 2010; 10:411.

(406) Althoff KN, Gange SJ, Klein MB, Brooks JT, Hogg RS, Bosch RJ et al. Late presentation for human immunodeficiency virus care in the United States and Canada. *Clin Infect Dis* 2010; 50(11):1512-1520.

(407) Wong KH, Lee SS, Low KH, Wan WY. Temporal trend and factors associated with late HIV diagnosis in Hong Kong, a low HIV prevalence locality. *AIDS Patient Care & Stds* 2003; 17(9):461-469.

(408) Delpierre C, Cuzin L, Lauwers-Cances V, Marchou B, Lang T. High-Risk groups for late diagnosis of HIV infection: a need for rethinking testing policy in the general population. *AIDS Patient Care STDS* 2006; 20(12):838-847.

(409) Borghi V, Girardi E, Bellelli S, Angeletti C, Mussini C, Porter K et al. Late presenters in an HIV surveillance system in Italy during the period 1992-2006. *J Acquir Immune Defic Syndr* 2008; 49(3):282-286.

(410) Kivela PS, Krol A, Salminen MO, Ristola MA. Determinants of late HIV diagnosis among different transmission groups in Finland from 1985 to 2005. *HIV Med* 2010; 11(6):360-367.

(411) Guy RJ, McDonald AM, Bartlett MJ, Murray JC, Giele CM, Davey TM et al. Characteristics of HIV diagnoses in Australia, 1993-2006. *Sex Health* 2008; 5(2):91-96.

(412) Delpierre C, Lauwers-Cances V, Pugliese P, Poizot-Martin I, Billaud E, Duvivier C et al. Characteristics trends, mortality and morbidity in persons newly diagnosed HIV positive during the last decade: the profile of new HIV diagnosed people. *Eur J Public Health* 2008; 18(3):345-347.

(413) Keruly JC, Moore RD. Immune status at presentation to care did not improve among antiretroviral-naive persons from 1990 to 2006. *Clin Infect Dis* 2007; 45(10):1369-1374.

(414) Mugavero MJ, Lin HY, Allison JJ, Willig JH, Chang PW, Marler M et al. Failure to establish HIV care: characterizing the "no show" phenomenon. *Clin Infect Dis* 2007; 45(1):127-130.

(415) Boyd AE, Murad S, O'Shea S, de Ruiter A, Watson C, Easterbrook PJ. Ethnic differences in stage of presentation of adults newly diagnosed with HIV-1 infection in south London. *HIV Med* 2005; 6(2):59-65.

(416) Jean-Jacques M, Walensky RP, Aaronson WH, Chang Y, Freedberg KA. Late diagnosis of HIV infection at two academic medical centers: 1994-2004. *AIDS Care* 2008; 20(8):977-983.

(417) Grangeiro A, Escuder MM, Menezes PR, Alencar R, Ayres de CE. Late entry into HIV care: estimated impact on AIDS mortality rates in Brazil, 2003-2006. *PLoS ONE* 2011; 6(1):e14585.

(418) Lucas SB, Curtis H, Johnson MA. National review of deaths among HIV-infected adults. *Clin Med* 2008; 8(3):250-252.

(419) Rodger AJ, Curtis H, Sabin C, Johnson M. Assessment of hospitalizations among HIV patients in the UK: a national cross-sectional survey. *Int J STD AIDS* 2010; 21(11):752-754.

(420) Gras L, van SA, Bezemer D, Smit C, Wit F, De WF. Lower mortality and earlier start of combination antiretroviral therapy in patients tested repeatedly for HIV than in those with a positive first test. *AIDS* 2011.

(421) Losina E, Schackman BR, Sadownik SN, Gebo KA, Walensky RP, Chiosi JJ et al. Racial and sex disparities in life expectancy losses among HIV-infected persons in the united states: impact of risk behavior, late initiation, and early discontinuation of antiretroviral therapy. *Clin Infect Dis* 2009; 49(10):1570-1578.

(422) Mugavero MJ, Lin HY, Willig JH, Westfall AO, Ulett KB, Routman JS et al. Missed visits and mortality among patients establishing initial outpatient HIV treatment. *Clin Infect Dis* 2009; 48(2):248-256.

(423) Klein D, Hurley LB, Merrill D, Quesenberry CP, Jr., Consortium for HIV/AIDS Interregional Research. Review of medical encounters in the 5 years before a diagnosis of HIV-1 infection: implications for early detection. *Journal of Acquired Immune Deficiency Syndromes: JAIDS* 2003; 32(2):143-152.

(424) Phair JP, Mellors JW, Detels R, Margolick JB, Munoz A. Virologic and immunologic values allowing safe deferral of antiretroviral therapy. *AIDS* 2002; 16(18):2455-2459.

(425) Schwartlander B, Horsburgh CR, Jr., Hamouda O, Skarabis H, Koch MA. Changes in the spectrum of AIDS-defining conditions and decrease

in CD4+ lymphocyte counts at AIDS manifestation in Germany from 1986 to 1991. *AIDS* 1992; 6(4):413-420.

(426) Ahdieh L, Gange SJ, Greenblatt R, Minkoff H, Anastos K, Young M 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):923-933.

(427) Emery S, Neuhaus JA, Phillips AN, Babiker A, Cohen CJ, Gatell JM et al. Major clinical outcomes in antiretroviral therapy (ART)-naive participants and in those not receiving ART at baseline in the SMART study. *J Infect Dis* 2008; 197(8):1133-1144.

(428) Phillips AN, Gazzard B, Gilson R, Easterbrook P, Johnson M, Walsh J et al. Rate of AIDS diseases or death in HIV-infected antiretroviral therapy-naive individuals with high CD4 cell count. *AIDS* 2007; 21(13):1717-1721.

(429) Opravil M, Ledergerber B, Furrer H, Hirschel B, Imhof A, Gallant S 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):1371-1381.

(430) Branson BM, Handsfield HH, Lampe MA, Janssen RS, Taylor AW, Lyss SB et al. Revised recommendations for HIV testing of adults, adolescents, and pregnant women in health-care settings. *MMWR Recomm Rep* 2006; 55(RR-14):1-17.

(431) Little RJ, Vartivarian S. On weighting the rates in non-response weights. *Stat Med* 2003; 22(9):1589-1599.

(432) Weinhardt LS, Carey MP, Johnson BT, Bickham MS. Effects of HIV counseling and Testing on Sexual Risk Behaviour: A meta-Analytic Review of Published Research, 1985-1997. *Am J Public Health* 1999; 89:1397-1405.

(433) Leaity S, Sherr L, Wells H, Evans A, Miller R, Johnson M et al. Repeat HIV testing: high-risk behaviour or risk reduction strategy? *AIDS* 2000; 14(5):547-552.

(434) Department of Health. The national strategy for sexual health and HIV. 1-53. 2001. Crown.

(435) Public Health Laboratory Service Communicable Disease Surveillance Centre, ICH (London), SCIEH. HIV & AIDS in the United Kingdom 2001: An update November 2002. London: Public Health Laboratory Service; 2002.

(436) Rapatski BL, Suppe F, Yorke JA. HIV epidemics driven by late disease stage transmission. *J Acquir Immune Defic Syndr* 2005; 38(3):241-253.

(437) Porco TC, Martin JN, Page-Shafer KA, Cheng A, Charlebois E, Grant RM et al. Decline in HIV infectivity following the introduction of highly active antiretroviral therapy. *AIDS* 2004; 18(1):81-88.

(438) Vernazza PL, Troiani L, Flepp MJ, Cone RW, Schock J, Roth F et al. Potent antiretroviral treatment of HIV-infection results in suppression of the seminal shedding of HIV. The Swiss HIV Cohort Study. *AIDS,* 2000; 14(2):117-121.

(439) Advancing HIV prevention: new strategies for a changing epidemic--United States, 2003. *MMWR Morb Mortal Wkly Rep* 2003; 52(15):329-332.

(440) Health Protection Agency, National Public Health Service for Wales, SCIEH, ISD, CDSC Northern Ireland and the UASSG. Focus on Prevention. HIV and other Sexually Transmitted Infections in the United Kingdom in 2003. An update: November 2004. London: Health Protection Agency; 2004.

(441) Burns FM, Johnson AM, Nazroo J, Ainsworth J, Anderson J, Fakoya A et al. Missed opportunities for earlier HIV diagnosis within primary and secondary healthcare settings in the UK. *AIDS* 2008; 22(1):115-122.

(442) Liddicoat RV, Horton NJ, Urban R, Maier E, Christiansen D, Samet JH. Assessing missed opportunities for HIV testing in medical settings. *J Gen Intern Med* 2004; 19(4):349-356.

(443) McDonald EA, Currie MJ, Bowden FJ. Delayed diagnosis of HIV: missed opportunities and triggers for testing in the Australian Capital Territory. *Sex Health* 2006; 3(4):291-295.

(444) Sullivan AK, Curtis H, Sabin CA, Johnson MA. Newly diagnosed HIV infections: review in UK and Ireland. *BMJ* 2005; 330(7503):1301-1302.

(445) Sabin CA, Smith CJ, Gumley H, Murphy G, Lampe FC, Phillips AN et al. Late presenters in the era of highly active antiretroviral therapy: uptake of and responses to antiretroviral therapy. *AIDS* 2004; 18(16):2145-2151.

(446) British Association for Sexual Health and HIV. Recommendations from the Bacterial Special Interest Group of BASHH: testing for sexually transmitted infections in primary care settings. 2004.

(447) Rogstad KE, Henton L. General practitioners and the National Strategy on Sexual Health and HIV. *Int J STD AIDS* 2004; 15(3):169-172.

(448) Deuffic-Burban S, Losina E, Wang B, Gabillard D, Messou E, Divi N et al. Estimates of opportunistic infection incidence or death within specific CD4 strata in HIV-infected patients in Abidjan, Cote d'Ivoire: impact of

alternative methods of CD4 count modelling. *Eur J Epidemiol* 2007; 22(10):737-744.

(449) Ray M, Logan R, Sterne JA, Hernadez-Diaz S, Robins JM, Sabin C et al. The effect of combined antiretroviral therapy on the overall mortality of HIV-infected individuals. *AIDS* 2010; 24(1):123-137.

(450) Sterne JA, Hernan MA, Ledergerber B, Tilling K, Weber R, Sendi P et al. Long-term effectiveness of potent antiretroviral therapy in preventing AIDS and death: a prospective cohort study. *Lancet* 2005; 366(9483):378-384.

(451) Mocroft A, Vella S, Benfield TL, Chiesi A, Miller V, Gargalianos P et al. Changing patterns of mortality across Europe in patients infected with HIV-1. EuroSIDA Study Group. *Lancet* 1998; 352(9142):1725-1730.

(452) van Sighem AI, Gras LA, Reiss P, Brinkman K, De WF. Life expectancy of recently diagnosed asymptomatic HIV-infected patients approaches that of uninfected individuals. *AIDS* 2010; 24(10):1527-1535.

(453) Zwahlen M, Harris R, May M, Hogg R, Costagliola D, De WF et al. Mortality of HIV-infected patients starting potent antiretroviral therapy: comparison with the general population in nine industrialized countries. *Int J Epidemiol* 2009; 38(6):1624-1633.

(454) Sullivan PS, Denniston M, McNaghten A, Buskin SE, Broyles ST, Mokotoff ED. Use of a population-based survey to determine incidence of AIDS-defining opportunistic illnesses among HIV-positive persons receiving medical care in the United States. *AIDS Res Ther* 2007; 4:17.

(455) Survival and progression of HIV disease in women attending GUM/HIV clinics in Britain and Ireland. Study Group for the MRC Collaborative Study of HIV Infection in Women. *Sex Transm Infect* 1999; 75(4):247-252.

(456) French CE, Antoine D, Gelb D, Jones JA, Gilbert RL, Watson JM. Tuberculosis in non-UK-born persons, England and Wales, 2001-2003. *Int J Tuberc Lung Dis* 2007; 11(5):577-584.

(457) Babiker A, Darbyshire J, Pezzotti P, Porter K, Rezza G, Walker SA 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):951-958.

(458) Perez-Hoyos S, Del Amo J, Muga R, del Romero J, Garcia dO, Guerrero R et al. Effectiveness of highly active antiretroviral therapy in Spanish cohorts of HIV seroconverters: differences by transmission category. *AIDS* 2003; 17(3):353-359.

(459) Chene G, Sterne JA, May M, Costagliola D, Ledergerber B, Phillips AN et al. Prognostic importance of initial response in HIV-1 infected patients starting potent antiretroviral therapy: analysis of prospective studies. *Lancet* 2003; 362(9385):679-686.

(460) Poundstone KE, Chaisson RE, Moore RD. Differences in HIV disease progression by injection drug use and by sex in the era of highly active antiretroviral therapy. *AIDS* 2001; 15(9):1115-1123.

(461) Effective therapy has altered the spectrum of cause-specific mortality following HIV seroconversion. *AIDS* 2006; 20(5):741-749.

(462) Mocroft A, Brettle R, Kirk O, Blaxhult A, Parkin JM, Antunes F et al. Changes in the cause of death among HIV positive subjects across Europe: results from the EuroSIDA study. *AIDS* 2002; 16(12):1663-1671.

(463) Mocroft A, Gatell J, Reiss P, Ledergerber B, Kirk O, Vella S et al. Causes of death in HIV infection: the key determinant to define the clinical response to anti-HIV therapy. *AIDS* 2004; 18(17):2333-2337.

(464) Smit C, Geskus R, Walker S, Sabin CA, Coutinho R, Porter K et al. Effective therapy has altered the spectrum of cause-specific mortality following HIV seroconversion. *AIDS* 2006; 20(5):741-749.

(465) Sabin CA, Smith CJ, Youle M, Lampe FC, Bell DR, Puradiredja D et al. Deaths in the era of HAART: contribution of late presentation, treatment exposure, resistance and abnormal laboratory markers. *AIDS* 2006; 20(1):67-71.

(466) Krentz HB, Kliewer G, Gill MJ. Changing mortality rates and causes of death for HIV-infected individuals living in Southern Alberta, Canada from 1984 to 2003. *HIV Med* 2005; 6(2):99-106.

(467) Crum NF, Riffenburgh RH, Wegner S, Agan BK, Tasker SA, Spooner KM 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):194-200.

(468) Palella FJ, Jr., Baker RK, Moorman AC, Chmiel JS, Wood KC, Brooks JT 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):27-34.

(469) Yang CH, Huang YF, Hsiao CF, Yeh YL, Liou HR, Hung CC et al. Trends of mortality and causes of death among HIV-infected patients in Taiwan, 1984-2005. *HIV Med* 2008; 9(7):535-543.

(470) Buchacz K, Baker RK, Palella FJ, Jr., Chmiel JS, Lichtenstein KA, Novak RM et al. AIDS-defining opportunistic illnesses in US patients, 1994-2007: a cohort study. *AIDS* 2010; 24(10):1549-1559.

(471) Smith C, Sabin CA, Lundgren JD, Thiebaut R, Weber R, Law M et al. Factors associated with specific causes of death amongst HIV-positive individuals in the D:A:D Study. *AIDS* 2010; 24(10):1537-1548.

(472) Del Amo J, Perez-Hoyos S, Moreno A, Quintana M, Ruiz I, Cisneros JM et al. Trends in AIDS and mortality in HIV-infected subjects with hemophilia from 1985 to 2003: the competing risks for death between AIDS and liver disease. *J Acquir Immune Defic Syndr* 2006; 41(5):624-631.

(473) Palella FJ, Jr., Delaney KM, Moorman AC, Loveless MO, Fuhrer J, Satten GA et al. Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. HIV Outpatient Study Investigators. *New England Journal of Medicine* 1998; 338(13):853-860.

(474) d'Arminio MA, Sabin CA, Phillips A, Sterne J, May M, Justice A et al. The changing incidence of AIDS events in patients receiving highly active antiretroviral therapy. *Arch Intern Med* 2005; 165(4):416-423.

(475) Chene G, Binquet C, Moreau JF, Neau D, Pellegrin I, Malvy D et al. Changes in CD4+ cell count and the risk of opportunistic infection or death after highly active antiretroviral treatment. Groupe d'Epidemiologie Clinique du SIDA en Aquitaine. *AIDS* 1998; 12(17):2313-2320.

(476) Lodwick RK, Sabin CA, Porter K, Ledergerber B, van SA, Cozzi-Lepri A et al. Death rates in HIV-positive antiretroviral-naive patients with CD4 count greater than 350 cells per microL in Europe and North America: a pooled cohort observational study. *Lancet* 2010; 376(9738):340-345.

(477) Miller V, Sabin CA, Phillips AN, Rottmann C, Rabenau H, Weidmann E et al. The impact of protease inhibitor-containing highly active antiretroviral therapy on progression of HIV disease and its relationship to CD4 and viral load. *AIDS* 2000; 14(14):2129-2136.

(478) Miller V, Phillips AN, Clotet B, Mocroft A, Ledergerber B, Kirk O et al. Association of virus load, CD4 cell count, and treatment with clinical progression in human immunodeficiency virus-infected patients with very low CD4 cell counts. *Journal of Infectious Diseases* 2002; 186(2):189-197.

(479) Ferreira OC, Jr., Suleiman G, Brites C, Novoa P, Piovesana M, Suleiman J et al. A rapid and reliable assay to enumerate CD4+ T lymphocytes in whole blood. *Journal of Acquired Immune Deficiency Syndromes* 1994; 7(12):1224-1227.

(480) Jacobson LP, Li R, Phair J, Margolick JB, Rinaldo CR, Detels R et al. Evaluation of the effectiveness of highly active antiretroviral therapy in persons with human immunodeficiency virus using biomarker-based equivalence of disease progression. *Am J Epidemiol* 2002; 155(8):760-770.

(481) Olsen CH, Gatell J, Ledergerber B, Katlama C, Friis-Moller N, Weber J et al. Risk of AIDS and death at given HIV-RNA and CD4 cell count, in relation to specific antiretroviral drugs in the regimen. *AIDS* 2005; 19(3):319-330.

(482) Mocroft A, Reiss P, Gasiorowski J, Ledergerber B, Kowalska J, Chiesi A et al. Serious fatal and nonfatal non-AIDS-defining illnesses in Europe. *J Acquir Immune Defic Syndr* 2010; 55(2):262-270.

(483) Keiser O, Spycher B, Rauch A, Calmy A, Cavassini M, Glass TR et al. Outcomes of Antiretroviral Therapy in the Swiss HIV Cohort Study: Latent Class Analysis. *AIDS Behav* 2011.

(484) 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):1387-1396.

(485) Mussini C, Cossarizza A, Sabin C, Babiker A, De LA, Bucher HC et al. Decline of CD4+ T cell count before start of therapy and immunological response to treatment in antiretroviral-naive individuals. *AIDS* 2011.

(486) Hughes R, Sterne J, Walsh J, Bansi L, Gilson R, Orkin C et al. Long-term trends in CD4 cell counts and impact of viral failure in individuals starting antiretroviral therapy: UK Collaborative HIV Cohort (CHIC) study. *HIV Med* 2011.

(487) Xiuhong L, MJ B, JB D, RC R, PJ P, JL P. CD4+ T-Cell Counts and Plasma HIV-1 RNA Levels Beyond 5 Years of Highly Active Antiretroviral Therapy (HAART). *J Acquir Immune Defic Syndr* 2011.

(488) Althoff KN, Justice AC, Gange SJ, Deeks SG, Saag MS, Silverberg MJ et al. Virologic and immunologic response to HAART, by age and regimen class. *AIDS* 2010; 24(16):2469-2479.

(489) French AL, Gawel SH, Hershow R, Benning L, Hessol NA, Levine AM et al. Trends in mortality and causes of death among women with HIV in the United States: a 10-year study. *J Acquir Immune Defic Syndr* 2009; 51(4):399-406.

(490) Lima VD, Lepik KJ, Zhang W, Muldoon KA, Hogg RS, Montaner JS. Regional and temporal changes in HIV-related mortality in British Columbia, 1987-2006. *Can J Public Health* 2010; 101(5):415-419.

(491) Prosperi MC, Cozzi-Lepri A, Antinori A, Cassola G, Torti C, Ursitti MA et al. Favourable evolution of virological and immunological profiles in treated and untreated patients in Italy in the period 1998-2008. *HIV Med* 2011; 12(3):174-182.

(492) Kaldor JM, Delpech V, Guy RJ. AIDS case reporting: do we still need it? *Lancet* 2008.

(493) Munoz A, Gange SJ, Jacobson LP. Distinguishing efficacy, individual effectiveness and population effectiveness of therapies. *AIDS* 2000; 14(6):754-756.

(494) Wolbers M, Babiker A, Sabin C, Young J, Dorrucci M, Chene G et al. Pretreatment CD4 cell slope and progression to AIDS or death in HIV-infected patients initiating antiretroviral therapy--the CASCADE collaboration: a collaboration of 23 cohort studies. *PLoS Med* 2010; 7(2):e1000239.

(495) Street E, Curtis H, Sabin CA, Monteiro EF, Johnson MA. British HIV Association (BHIVA) national cohort outcomes audit of patients commencing antiretrovirals from naive. *HIV Med* 2009; 10(6):337-342.

(496) Easterbrook PJ, Phillips AN, Hill T, Matthias R, Fisher M, Gazzard B et al. Patterns and predictors of the use of different antiretroviral drug regimens at treatment initiation in the UK. *HIV Med* 2008; 9(1):47-56.

(497) Stohr W, Dunn D, Porter K, Hill T, Gazzard B, Walsh J et al. CD4 cell count and initiation of antiretroviral therapy: trends in seven UK centres, 1997-2003. *HIV Med* 2007; 8(3):135-141.

(498) Smith CJ, Sabin CA, Youle MS, Kinloch-de Loes S, Lampe FC, Madge S et al. Factors influencing increases in CD4 cell counts of HIV-positive persons receiving long-term highly active antiretroviral therapy. *J Infect Dis* 2004; 190(10):1860-1868.

(499) Florence E, Lundgren J, Dreezen C, Fisher M, Kirk O, Blaxhult A et al. Factors associated with a reduced CD4 lymphocyte count response to HAART despite full viral suppression in the EuroSIDA study. *HIV Med* 2003; 4(3):255-262.

(500) Kaufmann GR, Bloch M, Zaunders JJ, Smith D, Cooper DA. Long-term immunological response in HIV-1-infected subjects receiving potent antiretroviral therapy. *AIDS* 2000; 14(8):959-969.

(501) Grabar S, Le M, V, Goujard C, Leport C, Kazatchkine MD, Costagliola D et al. Clinical outcome of patients with HIV-1 infection according to immunologic and virologic response after 6 months of highly active antiretroviral therapy. *Annals of internal medicine* 2000; 133(6):401-410.

(502) Piketty C, Weiss L, Thomas F, Mohamed AS, Belec L, Kazatchkine MD. Long-term clinical outcome of human immunodeficiency virus-infected patients with discordant immunologic and virologic responses to a protease inhibitor-containing regimen. *J Infect Dis* 2001; 183(9):1328-1335.

(503) Pakker NG, Kroon ED, Roos MT, Otto SA, Hall D, Wit FW et al. Immune restoration does not invariably occur following long-term HIV-1 suppression during antiretroviral therapy. INCAS Study Group. *AIDS* 1999; 13(2):203-212.

(504) Petrie A, Sabin C. Medical Statistics at a Glance. 3rd Edition ed. London: Wiley-Blackwell; 2009.

(505) British HIV Association guidelines for antiretroviral treatment of HIV seropositive individuals. BHIVA Guidelines Co-ordinating Committee. *Lancet* 1997; 349(9058):1086-1092.

(506) Stohr W, Dunn D, Porter K, Hill T, Gazzard B, Walsh J et al. CD4 cell count and initiation of antiretroviral therapy: trends in seven UK centres, 1997-2003. *HIV Med* 2007; 8(3):135-141.

(507) Curtis H. BHIVA Clinical Audit:Management of patients who switch therapy; re-audit of patients starting therapy from naïve. BHIVA; 2005

(508) Brettle RP, Raab GM, Ross A, Fielding KL, Gore SM, Bird AG. HIV infection in women: immunological markers and the influence of pregnancy. *AIDS* 1995; 9(10):1177-1184.

(509) Tuomala RE, Kalish LA, Zorilla C, Fox H, Shearer W, Landay A et al. Changes in total, CD4+, and CD8+ lymphocytes during pregnancy and 1 year postpartum in human immunodeficiency virus-infected women. The Women and Infants Transmission Study. *Obstet Gynecol* 1997; 89(6):967-974.

(510) van Benthem BH, Vernazza P, Coutinho RA, Prins M. The impact of pregnancy and menopause on CD4 lymphocyte counts in HIV-infected women. *AIDS* 2002; 16(6):919-924.

(511) Grabar S, Le M, V, Goujard C, Egger M, Leport C, Kazatchkine MD et al. Response to highly active antiretroviral therapy at 6 months and long-term disease progression in HIV-1 infection. *J Acquir Immune Defic Syndr* 2005; 39(3):284-292.

(512) Babiker A, Darbyshire J, Pezzotti P, Porter K, Prins M, Sabin CA et al. Short-term CD4 cell response after highly active antiretroviral therapy initiated at different times from seroconversion in 1,500 seroconverters. *J Acquir Immune Defic Syndr* 2003; 32(3):303-310.

(513) Gandhi RT, Spritzler J, Chan E, Asmuth DM, Rodriguez B, Merigan TC et al. Effect of baseline- and treatment-related factors on immunologic recovery after initiation of antiretroviral therapy in HIV-1-positive subjects: results from ACTG 384. *J Acquir Immune Defic Syndr* 2006; 42(4):426-434.

(514) Le Moing V, Thiebaut R, Chene G, Sobel A, Massip P, Collin F et al. Long-term evolution of CD4 count in patients with a plasma HIV RNA persistently <500 copies/mL during treatment with antiretroviral drugs. *HIV Med* 2007; 8(3):156-163.

(515) Bartlett JA, Fath MJ, DeMasi R, Hermes A, Quinn J, Mondou E et al. An updated systematic overview of triple combination therapy in antiretroviral-naive HIV-infected adults. *AIDS* 2006; 20(16):2051-2064.

(516) Wolbers M, Battegay M, Hirschel B, Furrer H, Cavassini M, Hasse B et al. CD4+ T-cell count increase in HIV-1-infected patients with suppressed viral load within 1 year after start of antiretroviral therapy. *Antivir Ther* 2007; 12(6):889-897.

(517) Mocroft A, Ledergerber B, Zilmer K, Kirk O, Hirschel B, Viard JP et al. Short-term clinical disease progression in HIV-1-positive patients taking combination antiretroviral therapy: the EuroSIDA risk-score. *AIDS* 2007; 21(14):1867-1875.

(518) Lapadula G, Torti C, Maggiolo F, Casari S, Suter F, Minoli L et al. Predictors of clinical progression among HIV-1-positive patients starting HAART with CD4+ T-cell counts > or =200 cells/mm3. *Antivir Ther* 2007; 12(6):941-947.

(519) Thiebaut R, Chene G, Jacqmin-Gadda H, Morlat P, Mercie P, Dupon M et al. Time-updated CD4+ T lymphocyte count and HIV RNA as major markers of disease progression in naive HIV-1-infected patients treated with a highly active antiretroviral therapy: the Aquitaine cohort, 1996-2001. *J Acquir Immune Defic Syndr* 2003; 33(3):380-386.

(520) Pakker NG, Notermans DW, De Boer RJ, Roos MT, de Wolf F, Hill A et al. Biphasic kinetics of peripheral blood T cells after triple combination therapy in HIV-1 infection: a composite of redistribution and proliferation. *Nat Med* 1998; 4(2):208-214.

(521) Collazos J, Asensi V, Carton JA. Sex differences in the clinical, immunological and virological parameters of HIV-infected patients treated with HAART. *AIDS* 2007; 21(7):835-843.

(522) Garcia F, de Lazzari E, Plana M, Castro P, Mestre G, Nomdedeu M et al. Long-term CD4+ T-cell response to highly active antiretroviral therapy according to baseline CD4+ T-cell count. *J Acquir Immune Defic Syndr* 2004; 36(2):702-713.

(523) Bosch RJ, Wang R, Vaida F, Lederman MM, Albrecht MA. Changes in the slope of the CD4 cell count increase after initiation of potent antiretroviral treatment. *J Acquir Immune Defic Syndr* 2006; 43(4):433-435.

(524) Micheloud D, Berenguer J, Bellon JM, Miralles P, Cosin J, Lopez-Bernaldo de Quiros JC et al. Negative influence of age on CD4(+) cell recovery after highly active antiretroviral therapy in naive HIV-1-infected patients with severe immunodeficiency. *J Infect* 2008; 56(2):130-136.

(525) Viard JP, Burgard M, Hubert JB, Aaron L, Rabian C, Pertuiset N et al. Impact of 5 years of maximally successful highly active antiretroviral therapy on CD4 cell count and HIV-1 DNA level. *AIDS* 2004; 18(1):45-49.

(526) Mocroft A, Phillips AN, Gatell J, Ledergerber B, Fisher M, Clumeck N et al. Normalisation of CD4 counts in patients with HIV-1 infection and maximum virological suppression who are taking combination antiretroviral therapy: an observational cohort study. *Lancet* 2007; 370(9585):407-413.

(527) Kaufmann D, Pantaleo G, Sudre P, Telenti A. CD4-cell count in HIV-1-infected individuals remaining viraemic with highly active antiretroviral therapy (HAART). Swiss HIV Cohort Study. *Lancet* 1998; 351(9104):723-724.

(528) Wood E, Hogg RS, Yip B, Tyndall MW, Sherlock CH, Harrigan RP et al. "Discordant" increases in CD4 cell count relative to plasma viral load in a closely followed cohort of patients initiating antiretroviral therapy. *J Acquir Immune Defic Syndr* 2002; 30(2):159-166.

(529) Kaufmann GR, Furrer H, Ledergerber B, Perrin L, Opravil M, Vernazza P et al. Characteristics, determinants, and clinical relevance of CD4 T cell recovery to <500 cells/microL in HIV type 1-infected individuals receiving potent antiretroviral therapy. *Clin Infect Dis* 2005; 41(3):361-372.

(530) Battegay M, Nuesch R, Hirschel B, Kaufmann GR. Immunological recovery and antiretroviral therapy in HIV-1 infection. *Lancet Infect Dis* 2006; 6(5):280-287.

(531) Babiker A, Darbyshire J, Pezzotti P, Porter K, Prins M, Sabin CA et al. Short-term CD4 cell response after highly active antiretroviral therapy initiated at different times from seroconversion in 1,500 seroconverters. *J Acquir Immune Defic Syndr* 2003; 32(3):303-310.

(532) Khanna N, Opravil M, Furrer H, Cavassini M, Vernazza P, Bernasconi E et al. CD4+ T cell count recovery in HIV type 1-infected patients is independent of class of antiretroviral therapy. *Clin Infect Dis* 2008; 47(8):1093-1101.

(533) Kaufmann GR, Perrin L, Pantaleo G, Opravil M, Furrer H, Telenti A et al. CD4 T-lymphocyte recovery in individuals with advanced HIV-1 infection receiving potent antiretroviral therapy for 4 years: the Swiss HIV Cohort Study. *Arch Intern Med* 2003; 163(18):2187-2195.

(534) Casado JL, Perez-Elias MJ, Antela A, Sabido R, Marti-Belda P, Dronda F et al. Predictors of long-term response to protease inhibitor therapy in a cohort of HIV-infected patients. *AIDS* 1998; 12(11):F131-F135.

(535) Ekouevi DK, Inwoley A, Tonwe-Gold B, Danel C, Becquet R, Viho I et al. Variation of CD4 count and percentage during pregnancy and after delivery: implications for HAART initiation in resource-limited settings. *AIDS Res Hum Retroviruses* 2007; 23(12):1469-1474.

(536) Wilson IB, Landon BE, Ding L, Zaslavsky AM, Shapiro MF, Bozzette SA et al. A national study of the relationship of care site HIV specialization to early adoption of highly active antiretroviral therapy. *Med Care* 2005; 43(1):12-20.

(537) Stone VE, Mansourati FF, Poses RM, Mayer KH. Relation of physician specialty and HIV/AIDS experience to choice of guideline-recommended antiretroviral therapy. *J Gen Intern Med* 2001; 16(6):360-368.

(538) Schneider JA, Zhang Q, Auerbach A, Gonzales D, Kaboli P, Schnipper J et al. Do hospitalists or physicians with greater inpatient HIV experience improve HIV care in the era of highly active antiretroviral therapy? Results from a multicenter trial of academic hospitalists. *Clin Infect Dis* 2008; 46(7):1085-1092.

(539) Kitahata MM, Van Rompaey SE, Dillingham PW, Koepsell TD, Deyo RA, Dodge W et al. Primary care delivery is associated with greater physician experience and improved survival among persons with AIDS. *J Gen Intern Med* 2003; 18(2):95-103.

(540) Laine C, Markson LE, McKee LJ, Hauck WW, Fanning TR, Turner BJ. The relationship of clinic experience with advanced HIV and survival of women with AIDS. *AIDS* 1998; 12(4):417-424.

(541) Trotta MP, Cozzi-Lepri A, Ammassari A, Vecchiet J, Cassola G, Caramello P et al. Rate of CD4+ cell count increase over periods of viral load suppression: relationship with the number of previous virological failures. *Clin Infect Dis* 2010; 51(4):456-464.

(542) Lok JJ, Bosch RJ, Benson CA, Collier AC, Robbins GK, Shafer RW et al. Long-term increase in CD4+ T-cell counts during combination antiretroviral therapy for HIV-1 infection. *AIDS* 2010; 24(12):1867-1876.

(543) Effect of baseline CD4 cell count on immunological response to antiretroviral therapy in therapy-naive HIV-1-infected individuals: analysis of the St Mary's HIV clinical database. Dept of Primary Care &

Population Sciences, Royal Free & University College Medical School, London, United Kingdom. 2008.

(544) Discordant CD4/RNA responses to HAART are strongly associated with high baseline CD4 count and low HIV RNA: analysis of 406 naive patients. 99 Sep 26; 1999.

(545) Walker AS, Peto TE, Babiker AG, Darbyshire JH. Markers of HIV infection in the Concorde trial. Concorde Co-ordinating Committee. *QJM* 1998; 91(6):423-438.

(546) Barnett AG, van der Pols JC, Dobson AJ. Regression to the mean: what it is and how to deal with it. *Int J Epidemiol* 2005; 34(1):215-220.

(547) Department of Health. National Health Service Act 2006: Chapter 251 - Control of Patient Information. 8-11-2006.